CHALMERS
UNIVERSITY OF TECHNOLOGY

Machine Learning for Optical Network Security Monitoring: A Practical Perspective

Citation for the original published paper (version of record):
Furdek Prekratic, M., Natalino Da Silva, C., Lipp, F. et al (2020). Machine Learning for Optical Network Security Monitoring: A Practical Perspective. Journal of Lightwave Technology, 38(11): 2860-2871. http://dx.doi.org/10.1109/JLT.2020.2987032

Machine Learning for Optical Network Security Monitoring: A Practical Perspective

Marija Furdek, *Senior Member, IEEE, OSA*, Carlos Natalino, *Member, IEEE,* Fabian Lipp, David Hock, Andrea Di Giglio, and Marco Schiano, *Senior Member, IEEE*

***Abstract*—In order to accomplish cost-efficient management of complex optical communication networks, operators are seeking automation of network diagnosis and management by means of Machine Learning (ML). To support these objectives, new functions are needed to enable cognitive, autonomous management of optical network security. This paper focuses on the challenges related to the performance of ML-based approaches for detection and localization of optical-layer attacks, and to their integration with standard Network Management Systems (NMSs).**

We propose a framework for cognitive security diagnostics that comprises an attack detection module with Supervised Learning (SL), Semi-Supervised Learning (SSL) and Unsupervised Learning (UL) approaches, and an attack localization module that deduces the location of a harmful connection and/or a breached link. The influence of false positives and false negatives is addressed by a newly proposed Window-based Attack Detection (WAD) approach. We provide practical implementation guidelines for the integration of the framework into the NMS and evaluate its performance in an experimental network testbed subjected to attacks, resulting with the largest optical-layer security experimental dataset reported to date.

***Index Terms*—optical network security, monitoring, machine learning, attack detection.**

I. Introduction

Optical networks, as the only viable technology for supporting the consistent network traffic growth, are critical communication infrastructure whose secure and reliable operation is fundamental for a myriad of overlay services and applications. The inherent vulnerabilities of optical network building blocks, i.e., optical fibers, amplifiers and switches, can be exploited to perform physical-layer attacks aimed at service disruption [2]. Such attacks can be performed by, e.g., gaining direct access to patch-panels or the fiber plant largely deployed beyond a secure perimeter.

Attack methods can be vastly different in their sophistication, damaging potential, and the difficulty of detecting and counteracting them. For example, cutting the fiber is a straightforward attack which affects all connections traversing the cut link, and is relatively easy to detect as it causes loss of signal at the receiver end. More sophisticated methods include insertion of harmful, jamming signals (in-band or out-of-band) upon tampering with the patch-panels or breaching the fiber, e.g. by abusing a well-known monitoring technique for creating temporary passive couplers by bending the fiber [3]. Damage from such attacks depends on the power and spectral properties of the jamming signal as well as the underlying optical network design. Optical layer can also be disrupted without necessarily breaching the fiber, e.g. by a polarization scrambling attack where the fiber is squeezed to cause fast-varying polarization state variations that result in errors [4]. The complexity of the effects physical-layer attack techniques have on optical channel parameters makes their detection a very challenging task.

In order to sustain the evolution towards flexible, programmable and autonomous systems, optical networks should support autonomous diagnostics and operation [5]. Integrated telemetry and network analytics [6] are critical for realizing the Observe-Analyze-Act control loop and improve network performance, cost efficiency, and security [7]. The recent proliferation of ML techniques into numerous aspects of optical networking has brought forth new and powerful methods for cognitive and automated management of optical network security. These techniques have been shown successful in, e.g., detecting unauthorized signals in the network [8] or identifying jamming and polarization scrambling attacks [4]. However, the practicality of such solutions and their integration into existing network management systems remains an open issue. Challenges related to the performance of ML-based tools for attack diagnostics include their accuracy and the granularity of information they can provide; the complexity of their training and inference; their scalability and robustness to changes in the network states and the overall security threat landscape; as well as the availability of representative and sufficient data. Challenges related to their integration with standard network management frameworks include the choice of extra software that needs to be integrated into the control cycle; the software architecture options available for their implementation; and the implications of these architectures to the monitoring system requirements in terms of capacity and communication overhead.

The goal of this paper is to provide guidelines on the implementation of ML-based framework for optical-layer security monitoring. The paper combines new research findings related

Manuscript received December 5, 2019, revised March 5, 2020, and accepted March 31, 2020. This article is based upon work from VR project "Safeguarding Optical Communication Networks from Cyber-Security Attacks", COST Action 15127 RECODIS and CELTIC-NEXT projects SENDATE-EXTEND and SENDATE-PLANETS. *(Corresponding author: Marija Furdek.)*

M. Furdek and C. Natalino are with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden (e-mail: {furdek, carlos.natalino}@chalmers.se).

F. Lipp and D. Hock are with Infosim GmbH & Co. KG, Würzburg, Germany (e-mail: {lipp, hock}@infosim.net).

A. Di Giglio and M. Schiano are with Telecom Italia, Turin, Italy (e-mail: {andrea.digiglio, marco.schiano}@telecomitalia.it).

A preliminary version of this work was presented at ECOC 2019 [1].

The implementation of this work is available at https://github.com/carlosnatalino/JLT-2020-ML-Practical-Perspective.

to the design, deployment and performance of the attack detection framework with practical insights gained from the first demonstration of ML-based physical-layer security monitoring in an optical network scenario in [1]. The contributions of this paper extend the work from [1] in several ways, summarized as follows:

- SSL algorithms, in addition to SL and UL approaches for detection of physical-layer attacks are evaluated on the largest experimental dataset reported to date,
- A novel Window-based Attack Detection (WAD) approach is proposed, based on an analytical method for assessing the impact of false positives and/or false negatives on the probability of raising a security alarm for a given observation window,
- An approach for formulating binary attack syndromes to localize the source of link-based, as well as connection-based attacks in the network is presented,
- An assessment of architectural decisions in terms of software and ML models is performed,
- An encompassing comparative performance analysis of the three ML techniques is carried out, and the impact of their performance on the overall attack detection is evaluated.

II. Background and Related Work

A. Autonomous optical network management

To cope with increasing network heterogeneity and dynamicity, operators are striving towards cognition-driven automation of network planning and management workflows, abandoning the unreliable, ineffective and unscalable use of pre-determined thresholds on network performance indicators as triggers for (re)configuration actions, and embedding intelligent analytic techniques to analyse root cause of faults [6].

A generic transport Software Defined Networking (SDN)-integrated architecture for cognitive assurance is described in [6], embedding an analytics integration engine for fault detection into a state-of-the-art SDN framework. The authors present different options of running the engine, i.e., as a third party application, transport application, or integrated with the orchestrator, the transport SDN, or the hypervisor, and describe performance and interoperability trade-offs incurred by these alternatives. Protocols and models enabling advanced real-time SDN telemetry services in optical networks are described in [9], along with an experimental demonstration of an on-demand streaming telemetry service that can run either embedded into the SDN control plane using NETCONF protocol over a dedicated connection, or independently using a different API, such as gRPC with better support for telemetry data streaming.

Autonomous network operation relies on closing the loop between collecting the telemetry data, analyzing this data by applying different examination and interpretation functions, and performing actions necessary to maintain high network performance. This loop is referred to as the Observe-Decide-Act loop in [10], Observe-Analyze-Act loop in [7], and Collect-Analyze-Test (CAT) loop in [5]. The work in [10] focuses on increasing the efficiency of network resource usage by estimating Quality of Transmission (QoT) and applying ML techniques for margin reduction, along with experimentally demonstrating dynamic connection provisioning and rate adaptation under fiber/amplifier and Reconfigurable Optical Add-Drop Multiplexer (ROADM) aging, as well as frequency correction. In [7], the authors investigate similar use cases and discuss key requirements, advantages and drawbacks of centralized, distributed and hierarchical monitoring and data analytics capabilities, indicating main issues to be addressed before the potential of these tools can be fully utilized. In [5], key requirements on network diagnosis are examined from an operator's perspective on use cases related to 5G, optical transport disaggregation and multioperator orchestration, detailing on the role of the CAT loop as an enabler of truly autonomous, programmable networks.

All of the above approaches rely on the widespread and effective deployment of a plethora of ML techniques. In [11], the authors elaborate on key factors that drive the proliferation of ML in optical network management, overview typical ML algorithms and their application to optical networking problems, and list major challenges in the techniques' widespread deployment. ML-enabling data management issues are shown to encompass the discrepancies in network data sources (e.g., event logs, sensors, probes, signalling), monitoring device heterogeneity (in terms of, e.g., interfaces and protocols), data storage and representation.

The potential of ML tools to support complex tasks of autonomous network management has been shown in numerous application scenarios. A comprehensive survey of artificial intelligence techniques applicable to optical networking, which include ML as their subset, can be found in [12]. ML techniques applied to optical communications, as outlined in [13], typically attempt to perform regression, i.e., provide a functional description of given data in an effort to predict future values, or classification, i.e., derive decision boundaries between different data classes.

In doing so, Supervised Learning (SL) learning techniques (e.g., Artificial Neural Network (ANN) or Support Vector Machine (SVM)) utilize the *a priori* knowledge of class labels during training. The data can be labeled through carefully designed experiments by the experimenters or through automatized labeling. Labels can also be obtained from the network management system based on previous occurrences of events of interest. SL models are appropriate when complete information regarding what should be learned is available, i.e., the dataset defines exactly the inputs given to and the outputs expected from the model. During training, the parameters of the model are adjusted to match the observed outputs to the expected ones as close as possible.

In Unsupervised Learning (UL) (e.g., K-means clustering or principal components analysis), such labels are not available, but the algorithm learns to identify similarities between different inputs to cluster the data or extract features. UL models are appropriate when the dataset has no clear input/output nor strictly defined normal/abnormal conditions, but should be grouped by similarity, conversely substantially separating diverging points based on the intuition that anomalous samples are much rarer than the normal ones. UL usually has no

training phase, and the entire dataset needs to be traversed whenever new data points are included.

Semi-Supervised Learning (SSL) techniques (e.g., one-class support vector machine) combine the two approaches above, and are applicable in cases when labelling the majority of the data is too costly or infeasible, so only a small subset of the data is labeled [14]. For instance, when applied to anomaly detection, the training dataset should contain the normal data, and the algorithm should detect new data that is significantly different from that in the training dataset. During training, the parameters of the model are adjusted to enclose the normal samples with a spatial region as tightly as possible, facilitating the detection of data that falls outside the region as anomalous.

Approaches from the above categories have been applied to various optical networking tasks, such as performance prediction and fault diagnosis. Optical path performance prediction is typically addressed with different SL models, as in [15], where the contribution of different features to the regression error is analysed, [16], where a detailed comparison to existing theoretical models is performed, and [17], where the impact of the new lightpath to existing network connections is also taken into account.

Cognitive detection of so-called soft failures that result in gradual performance degradation of lightpaths is also often addressed with SL techniques. Component power and temperature are used to predict transceiver board failures in [18]. Detection and identification of failures caused by filter misalignment and undesired amplifier gain reduction is performed in [19] by analysing the Bit Error Rate (BER). Localization of filter shift and tightening is investigated in [20] by applying multi-classifier, single-classifier and residual computation approaches on the signal spectrum. In [21], the authors combine UL with SL for single-point and end-to-end detection of anomalies generated by excessive, varying signal attenuation. The data is not labeled beforehand, but UL performs density-based clustering to analyze patterns in the monitored data which are then further analyzed with an SL module. An encompassing tutorial on ML for failure management is presented in [22].

B. Optical network security management

Security is an essential aspect of truly autonomous network operation capable of diagnosing and counteracting breaches. Management of optical network security relies on three pillars: prevention, detection and reaction to attacks. Prevention of attacks encompasses risk modeling, vulnerability assessment and minimization of the attack surface through attack-aware network design and operation. Overviews of physical-layer vulnerabilities in evolving optical networks and attack techniques that exploit them to disrupt services can be found in [2], [23]. Awareness of physical-layer security was first introduced to optical network design, i.e. connection routing, in [24]. Approaches for attack-aware routing and spectrum allocation have been developed for static [25], periodic [26] and dynamic traffic [27].

Detection of attacks, which is in the focus of this paper, entails continuous monitoring of optical channels' performance, correctly attributing the observed degradation to a security breach, and determining the source location. Some attack techniques, such as high-power jamming, or tapping, can be detected by tracking the associated power surges or power drops, respectively. An approach for detecting intrusion-triggered power losses in passive optical network was proposed in [28]. Algorithms in [29], [30] localize the source of high-power jamming attacks by detecting power surges and comparing the power levels at the input and the output side of each node to detect the most upstream node where this relationship is abnormal. Many Optical Performance Monitoring (OPM) techniques, whose overview can be found in [31], require specialized devices such as Optical Spectrum Analyzers (OSAs). An OSA-based approach for detecting intrusion signals can be found in [8]. However, due to their prohibitive cost, OSAs are typically deployed only at a limited number of network sites. The advances in Digital Signal Processing (DSP)-enabled coherent transceivers allow the NMS to obtain a rich OPM dataset which, combined with data analytic tools, can be used for security diagnostic purposes without the need for costly monitoring devices. Determining the location of attacks which do not necessarily cause changes in the power levels was addressed in [32] by modeling the scope of harmful connections with binary words called Attack Syndromes (AttSyns). When AttSyns are unique for all attack scenarios, the harmful connection that caused the attack can be deduced from the subset of affected connections. Otherwise, additional security monitors or monitoring probes, whose resource-minimizing design was proposed in [33], are necessary to provide AttSyn disambiguation.

Mitigation of attacks encompasses fast and efficient recovery of the affected services as well as network adaptation to reduce the vulnerability to future attack occurrences. Protection from eavesdropping and jamming attacks via fast frequency hopping over fewest-shared-link multi-paths was proposed in [34]. Planning of protection resources capable of guaranteeing protection from jamming attacks based on identifying overlaps in attack scopes of the working and the backup path of each connection was presented in [35]. In [36], the authors experimentally demonstrated mitigation of attacks in Quantum Key Distribution (QKD)-enabled optical networks under physical-layer attacks aimed at key-rate degradation.

ML-based Attack Detection and Identification (ADI) functions are new features in the family of network analytics. While existing works consider connection degradation caused by component faults, management of physical-layer security threats remains an open issue. ML was applied to optical network security in [8], where the optical spectrum is scrutinized with SVM to identify malicious light sources traversing unauthorized paths. In [37], ANN was applied to detect high-power jamming and identify promising attack mitigation strategies. In [4], we applied various supervised learning approaches for detecting in-band, out-of-band jamming and polarization scrambling attacks on an optical link by analysing the experimental OPM data collected from a coherent receiver. In [38], techniques based on unsupervised learning were devised to detect attacks previously unseen and untrained for. First efforts towards integrating attack detection and localization into optical network management were demonstrated in [1].

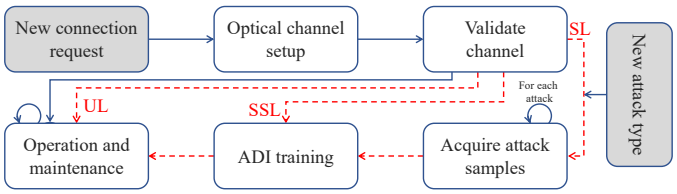


Fig. 1. Steps taken by SL, SSL and UL to accommodate for a new connection request or a newly discovered physical-layer attack type. Continuous lines represent the traditional process of establishing, operating and maintaining a connection. Dashed lines represent the steps inherent to the ML techniques.

In this paper, we extend upon the work in [1] by developing an encompassing ML-based framework for optical network security diagnostics, evaluating its performance on experimental data from an operator testbed, and providing guidelines on incorporating it into the NMS.

III. Network-Wide Attack Detection and Identification Approaches

Detection and identification of attacks can be performed by different ML techniques. Previous works have investigated the use of SL [4] and UL [38] from a single-link and single Optical Channel (OCh) perspective. However, adopting a network-wide multi-OCh Attack Detection and Identification (ADI) solution presents challenges and brings concerns beyond accuracy. Therefore, the benefits and drawbacks of each ML technique need to be assessed considering not only the accuracy of the models, but also their implications to the network operation.

The main differences between supervised, semi-supervised and unsupervised learning techniques are in their dataset requirements and training processes. Understanding how these models can be integrated into the NMS is of fundamental importance for efficient and reliable security assessment. Fig. 1 illustrates the steps inherent to SL, SSL and UL to accommodate for a new connection request or a newly discovered physical-layer attack type. A standard NMS action upon arrival of a new connection request is to set up a new optical channel, which may involve several channel estimation and resource assignment decisions. Once the channel is set up, a channel validation step verifies if the physical-layer conditions are appropriate, and, if yes, collects a set of OPM samples that are regarded as Normal Operating Conditions (NOC), i.e., represent the baseline OCh performance in attack-free conditions. The NOC samples are used for training in SSL and as baseline for SL. If SL is used, a representative dataset with channel performance in the presence of all known attack techniques must also be collected and labeled for ADI training purposes. This can incur complexities in designing and performing the experiments and may delay the operational phase of the OCh. For SSL, the attack-specific dataset collection step is bypassed and the model is retrained using only the NOC samples. UL skips both steps and proceeds with operation as before the newly arrived connection or attack type.

The main properties of ML techniques investigated in this paper are summarized Table I and highlighted as positive, negative and critical. All techniques require acquisition of NOC data and perform attack detection, i.e., classify OCh condition between normal and under-attack. Due to the labeling of attack samples, only SL models are able to perform attack identification, i.e., classify the specific attack technique (and its intensity) disrupting the OCh. The attack identification capability reflects on the need for a training dataset containing samples from all attack conditions, i.e., attack-specific data, resulting in a high training complexity observed for SL. Meanwhile, only NOC samples are required for training SSL models, resulting in a low training complexity. Both SL and SSL require re-training when new connections are established, given that they may traverse different network components than any existing channel. In this regard, UL does not require any (re-)training.

TABLE I
Summary of the Benefits and Drawbacks of Supervised Learning (SL), Semi-Supervised Learning (SSL) and Unsupervised Learning (UL) for ADI.

Property	SL	SSL	UL
Requires NOC data	Yes	Yes	Yes
Attack detection	Yes	Yes	Yes
Attack identification	Yes	No	No
Requires attack-specific labeled data	Yes	No	No
Training complexity	High	Low	None
Re-training for new OChs	Yes	Yes	No
Inference complexity	Low	Low	High
Requires prior samples	No	No	Yes
Supports stateless operation	Yes	Yes	Yes

NOC: Normal Operating Conditions; OCh: Optical Channel; green: positive; yellow: negative; red: critical.

Inference complexity follows an opposite trend from the training complexity. SL and SSL perform inference efficiently due to the fact that they have parameters that store the learned properties extracted from the dataset during training. On the other hand, UL needs to traverse the entire dataset every time a new prediction is made, which affects its efficiency and scalability. In the UL case, this dataset is usually composed of a number of prior samples that are used to characterize the NOC. Finally, due to not requiring prior samples, SL and SSL models have better support for stateless operation (which is discussed in more detail in Sec. V).

In order to evaluate performance of an ADI model, several metrics should be acquired. A fundamental performance indicator refers to the correctness in identifying the analyzed OPM samples. The samples obtained during an attack that are correctly classified as attacks are referred to as true positives, while their opposite, those not classified as attacks are false negatives. The NOC samples, i.e., those obtained in the absence of attack that are correctly classified as attack-free are true negatives, while their opposite, those wrongly classified as attacks are false positives. The true positive rate is denoted by T_P, the false negative by F_N, the true negative by T_N, and the false positive rate by F_P.

In this paper, we use the $f1$ *score* to evaluate the accuracy of the ML models. The best ML model, or a particular model configuration, can be selected according to the highest $f1$ score. The $f1$ score is defined in (1) a function of the precision (P) and recall (R). Precision (P) is defined in (2) and measures the sensitivity of an ML model to false positives. Recall

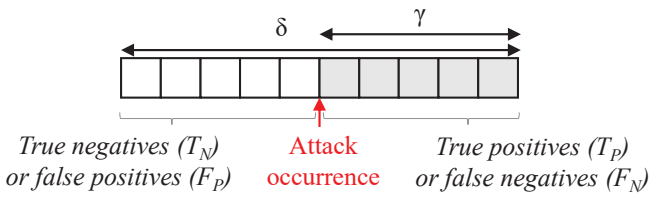


Fig. 2. Possible occurrences of false positive and false negative samples in an observation window δ.

(R) measures the sensitivity of a model to false negatives, as defined in (3). The $f1$ score balances the importance of precision and recall into a single metric, which is equal to 1 when the model is perfectly accurate in detecting attacks, and smaller otherwise.

$$f1 = 2 \times \frac{P \times R}{P + R} \tag{1}$$

$$P = \frac{T_P}{T_P + F_P} \tag{2}$$

$$R = \frac{T_P}{T_P + F_N} \tag{3}$$

A. Window-based Attack Detection (WAD)

To address the issue of false positive and false negative rates of the ML algorithms and reduce the impact of possible fast oscillations in the detected security status, we propose a Window-based Attack Detection (WAD) approach for security diagnostics. Instead of directly raising security alarms based on the ML output, WAD applies an additional inquiry mechanism by defining an observation window and setting a threshold on the number of samples in which attack presence must be detected before an alarm is triggered. In order to set the observation period and the alarm threshold in a way which ensures high attack detection accuracy, it is important to gauge the impact of false positive and negative rates on the overall security diagnostic performance.

To analyse the impact of false positive and negative rates, let us use a simple example shown in Fig. 2. Let δ denote the total number of samples considered for security monitoring, while γ ($\gamma \leq \delta$) is the number of samples collected *after* the attack occurrence. Before the attack, the security status of the analyzed connection can either be reported correctly as attack-free (true negative) or incorrectly as affected by an attack (false positive). After the attack, the status can either be reported correctly as affected by an attack (true positive), or incorrectly as attack-free (false negative), as depicted in Fig. 2. Note that $T_N + F_P = 1$ holds over the $\delta - \gamma$ samples and $T_P + F_N = 1$ holds over the γ samples.

If the attack detection module requires a threshold of τ positives out of the observed δ samples in order to raise an alarm, the probability of detecting an attack upon its occurrence can be expressed as follows:

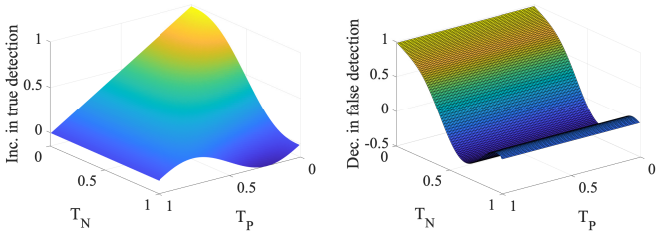


Fig. 3. Benefits obtained by the window-based approach for different true negative (T_N) and true positive (T_P) rates: (a) increase in true and (b) decrease in false attack detection (δ=100, γ=10 and τ=3).

$$\sum_{l=0}^{\tau} \binom{\delta-\gamma}{l} F_P^l T_N^{\delta-\gamma-l} \cdot \sum_{k=\tau-l}^{\gamma} \binom{\gamma}{k} T_P^k F_N^{\gamma-k} + \sum_{l=\tau+1}^{\delta-\gamma} \binom{\delta-\gamma}{l} F_P^l T_N^{\delta-\gamma-l} \cdot \sum_{k=0}^{\gamma} \binom{\gamma}{k} T_P^k F_N^{\gamma-k}. \tag{4}$$

The first product in (4) refers to the case when there are 0 to τ false positives prior to the attack and 0 to γ true positives upon the attack, such that there are τ or more positive (true or false) samples in total. The second product in (4) refers to the case when there are τ or more false positives prior to the attack and 0 to γ true positives upon the attack. In case there is no attack ($\gamma = 0$), an alarm can only be attributed to the false positives, and the probability of raising it is given by:

$$\sum_{l=\tau}^{\delta} \binom{\delta}{l} F_P^l T_N^{\delta-l} \tag{5}$$

Based on (4) and (5), Figs. 3a and b show the general benefits of the proposed window-based approach in terms of an increase in true detection and a decrease in false detection compared to directly raising alarms, respectively, as a function of the T_N and T_P values for fixed δ, γ and τ. A detailed analysis of the impact of δ, γ, and τ on the probability of a security alarm being raised by the WAD approach combined with the developed UL and SSL models based on their T_N and T_P performance is presented in Section VII.

IV. Attack Source Localization

For an encompassing physical-layer security monitoring framework, it is necessary not only to accurately detect when connections are affected by a physical-layer attack, but also to determine the location of the breach as a prerequisite of its neutralization. To this end, we define a framework for network-wide attack source localization based on identifying the scopes of different attack techniques and evaluating their effects for different physical locations of potential breaches.

The scope of an attack is modelled by binary words which we refer to as Attack Syndromes (AttSyns) [32]. The length of AttSyns matches the number of connections in the network and each bit denotes the security status of the corresponding connection recorded at the destination, where the regular,

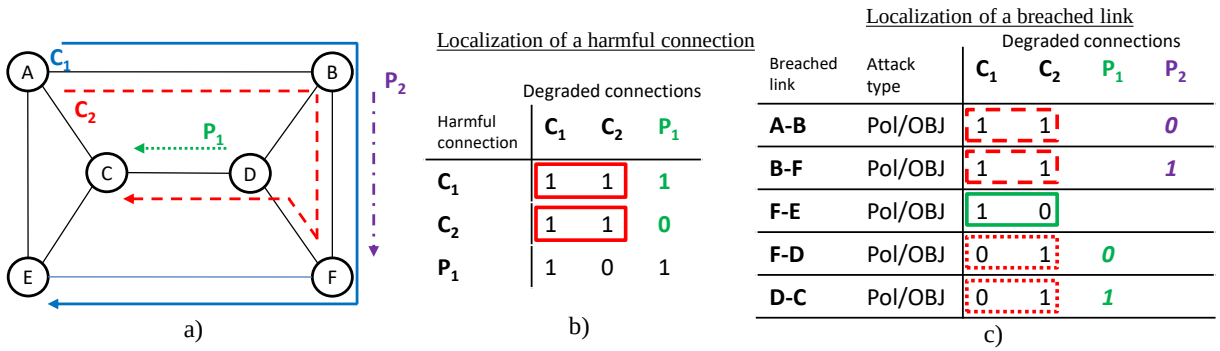


Fig. 4. Attack syndromes for a simple network example (a) to localize a harmful connection (b) or a breached link (c).

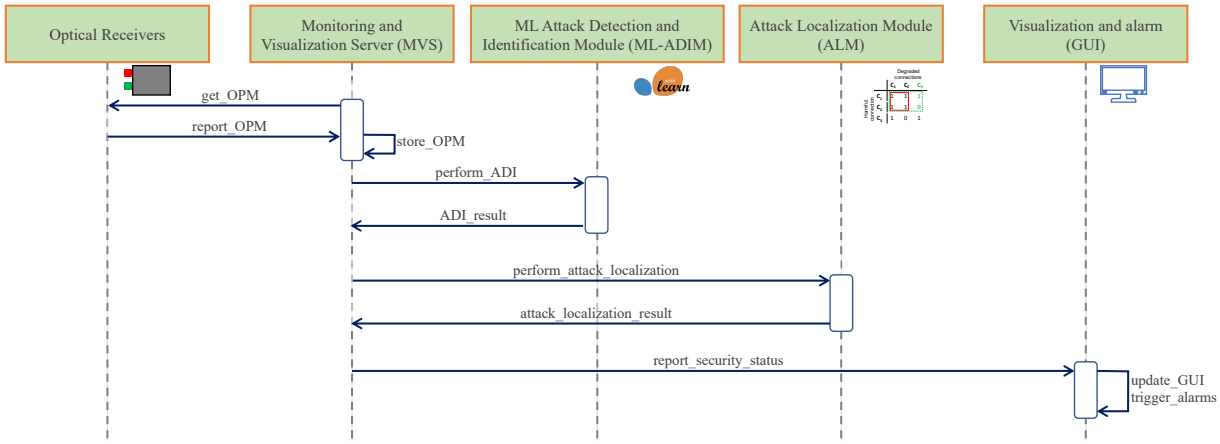


Fig. 5. Communication diagram among the modules for optical network security monitoring.

attack-free status is denoted by 0 while an affected status is denoted by 1. Attack syndromes are formed for each connection during setup by identifying the effects of the considered attack technique and determining whether the connection is affected for different potential locations of the attack. The AttSyns formed for different locations of a particular attack technique are stored and looked up for comparison with temporally evolving security statuses of connections.

If the attack syndromes for each attack scenario are unique, the source of an attack can be determined without additional diagnostic effort. If they are not unique, additional monitoring insights are needed. These can be obtained by adding monitoring probes or strategically placing OPM devices at intermediate nodes in the network. To avoid the extra cost of OPM devices, we rely on establishing monitoring probes to resolve ambiguities in AttSyns [32].

The AttSyn-based approach for attack source localization is illustrated on a simple example shown in Fig. 4a. Initially, only connections C_1 and C_2 are present in the network. If we consider an attack scenario where user connections registered in the system (i.e., C_1 or C_2 in the example) can become the source of an attack (e.g., by tampering with their power levels) and affect any other connections they share links with, the AttSyns are formed for each connection as the attack source, as shown in Fig.4b [32], [33]. In the AttSyn table, the rows match the attack source, while the columns correspond to the resulting degradation of each connection. If polarization scrambling (Pol) and out-of-band jamming (OBJ) attacks are considered, the model has to be adapted to the properties of these attacks. These attacks degrade all connections that traverse the breached link, so AttSyns need to be formed for each link as a possible source location, as shown in Fig. 4c.

As can be seen from the two AttSyn tables in Fig. 4, there is a substantial overlap of syndromes generated for the case when only C_1 and C_2 are established in the network. Adding P_1 as a monitoring probe resolves this ambiguity in the table in Fig. 4b. The table in Fig. 4c contains two pairs of matching syndromes, denoted with red dashed frames. To resolve the overlap, one discerning bit must be added to each syndrome in a matching pair, which is achieved by establishing monitoring probes P_1 and P_2.

The AttSyn tables can be defined and disambiguated for one specific attack methodology (e.g., with link-wise or connection-wise effects only) or over a larger set of attack techniques (e.g., with link- and connection-wise effects combined). The former is applicable when the attack localization module receives fine-granular attack identification information and can utilize the information on attack method to narrow down the search to the matching AttSyn table. The latter is needed when the attack identification module reports only the degradation of connections, without providing insight into the type of attack. In this case, the syndromes must be disambiguated across the entire AttSyn table, as is the case in the example shown in Fig. 4. Note that the AttSyn-

based intrusion localization does not apply to in-band jamming attacks, where the damaging effects remain confined to the affected connection, which requires monitoring at intermediate nodes.

V. ML-Aided Optical Security Monitoring

Fig. 5 illustrates the modules of the security diagnostic framework developed in this work, as well as their communication during one monitoring cycle. The Monitoring and Visualization Server (MVS) is the central monitoring agent, which is usually a standalone software specifically developed for network and service management, e.g., StableNet [39]. It is responsible for polling the optical transceivers using an appropriate protocol, e.g., NETCONF, and storing the reported OPM parameters in a long-term data storage for further processing. Once OPM data for the current cycle has been collected for all the active connections in the network, the MVS invokes the Machine-Learning-based Attack Detection and Identification Module (ML-ADIM), which uses one of the ML models explored in Sec. III to assess the security status of each connection. If the Window-based Attack Detection approach from Sec. III-A is adopted, the ML-ADIM will, in addition to the current OPM data, also require the previous $\delta - 1$ samples forming the observation window. WAD can be implemented as a separate module too, but for the sake of simplicity in the architecture and message exchange, we decided to integrate it into the ML-ADIM. Provided with the security statuses of all connections, the MVS invokes the Attack Localization Module (ALM), which deploys the appropriate techniques described in Sec. IV to determine the harmful connection or the compromised link. The reply from ALM is used by MVS to update its Graphical User Interface (GUI) and dispatch the outcome of the security diagnostic procedure to the network manager. One option to realize the communication between the MVS, ML-ADIM and ALM is to use Representational State Transfer (REST) services with persistent connections exchanging data using compressed JavaScript Object Notation (JSON) encoding, which results in an efficient use of network resources.

An important design decision for the modules in Fig. 5 is the network state management. If the *micro-service* and *stateless* principles of web-service design are followed, the modules for attack detection and localization should be as simple as possible. As a result, the MVS should include all inputs required by each module every time they are called. For the ML-ADIM, if UL is selected as the model of choice, prior samples (discussed in Table I) should be included. For the ALM, the pre-computed AttSyns should be included. While this option reduces the complexity of the modules, bringing significant benefits in terms of simpler deployment, easier migration and scaling according to the network load, it comes at the cost of increased traffic between the MVS and the other modules. Finally, since these modules will receive all the input they need from the MVS, the same instance of the service can serve multiple domains owned by the same entity.

If ML-ADIM and ALM are deployed as *stateful* services, each of these modules should contain the means to store the long-term information necessary for their execution. The ML-ADIM should store the prior samples (if necessary), and the ALM should store the topological, routing and AttSyn information. Both of these modifications would reduce the size of the messages exchanged with the MVS. However, the state management of stateful services makes them more resource-demanding (in terms of processing and storage), and more challenging to adapt to changing network conditions (i.e., to migrate, scale, or recover from faults). Due to the stateful property, these models are not appropriate for security assessment in multi-domain scenarios.

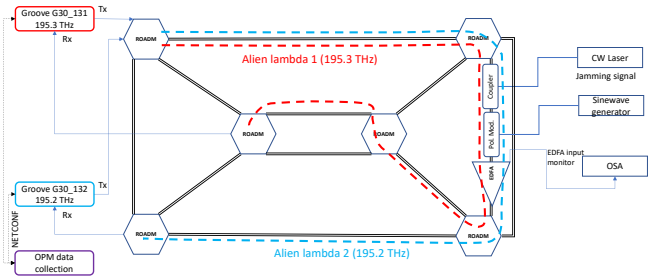


Fig. 6. Network testbed layout of the optical security experiment.

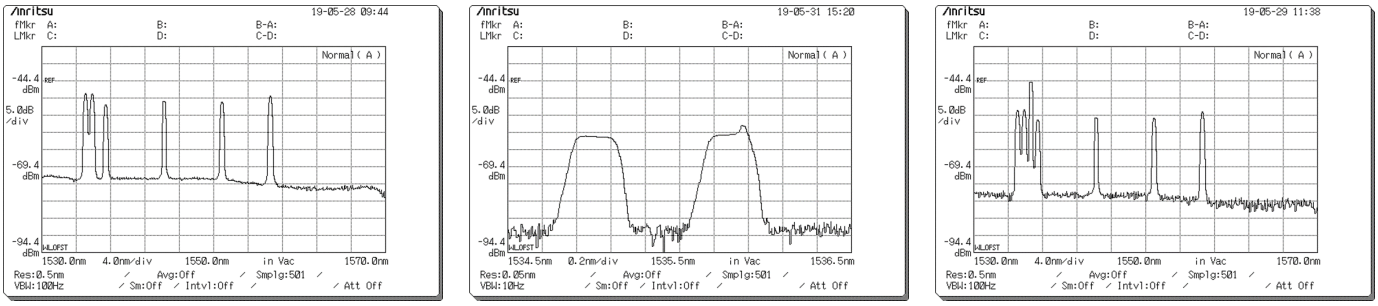
TABLE II
Optical Performance Monitoring (OPM) Parameters of Each Data Sample

Acronym	Description
CD	Chromatic Dispersion
DGD	Differential Group Delay
OSNR	Optical Signal to Noise Ratio
PDL	Polarization Dependent Loss
Q-factor	Q factor
BE-FEC	Block Errors before FEC
BER-FEC	Bit Error Rate before FEC
UBE-FEC	Uncorrected Block
BER-POST-FEC	Bit Error Rate after FEC
OPR	Optical Power Received
OPT	Optical Power Transmitted
OFT	Optical Frequency Transmitted
OFR	Optical Frequency Received
LOS	Loss Of Signal

For all parameters except BE-FEC, UBE-FEC, and LOS, the system provides the maximum, minimum and average values in the observation interval.

VI. Optical Jamming Experiments in a Network Testbed

The proposed ML-based framework for cognitive security assessment has been assessed in an experimental optical network testbed environment. The testbed, shown in Fig. 6, is based upon commercial optical transport network technology and is composed by 6 ROADM nodes, 1 EDFA amplification node and 10 links. The ROADMs are fixed-grid multidegree nodes capable of carrying 80 optical channels at 50 GHz spacing. The links are composed of 25 dB optical attenuators to emulate fibre loss. We have inserted a 3 dB fibre coupler (emulating a bent-fibre temporary coupler that might be used for a field attack) and a polarization modulator in the amplified link to enable the injection of a jamming signal and the high speed polarization modulation attack as described later.



(a) Unperturbed signal (b) In-band attack scheme (c) Out-of-band attack scheme

Fig. 7. (a) Power spectrum of the unperturbed DWDM signal at the EDFA node monitoring port. The OSA resolution is 0.5 nm. (b) Power spectrum of the 195.2 THz OCh under test in the in-band attack scheme (-7 dB intrusion signal power with respect to the OCh under attack). The 195.3 THz OCh is unaffected by this attack. (c) Power spectrum of all channels in the out-of-band attack scheme (+8.7 dB intrusion signal power with respect to the OCh under test). The intrusion signal is the one just at the right-hand side of the OCh under test.

The polarization modulator is an all-fiber component made by 3 piezoelectric fiber squeezers that produce stress-induced birefringence and hence polarization modulation. Just one of the fiber squeezers is used. It is driven by a sinewave signal at 136 kHz frequency, which corresponds to one of the resonant frequencies of the piezoelectric element and therefore produces a deep polarization modulation even with modest driving voltage. The polarization modulator loss is negligible, and it does not perturb the DWDM signal except when the sinewave generator is switched on to simulate a polarization attack. A Continuous Wave (CW) tunable laser is connected to the 3 dB coupler as a jamming signal generator.

The OChs under test are two 200 Gbit/s polarization multiplexed 16QAM signals generated by two commercial transponders (Infinera Groove G30). The two OChs, whose frequencies are 195.2 and 195.3 THz (1535.82 and 1535.04 nm wavelength respectively), are handled by the network as alien lambdas and are routed on two paths that share the amplified link where the attacks are implemented. The network is loaded with 4 additional coherent OChs spanning from 1537 nm to 1556 nm wavelength.

An OSA is connected to the monitor port of the EDFA node to measure the power spectra of the DWDM signal just after the injection of the jamming signal. The power spectrum of the unperturbed DWDM signal at the EDFA node monitoring port is shown in Fig. 7a. In normal working condition, i.e. without any intrusion signal or polarization modulation, the two OChs under test operate error-free with 32 dB Optical Signal-to-Noise Ratio $(\mathrm{OSNR})_{0.1}$ (measured with 0.5 nm resolution and rescaled to 0.1 nm).

The network management system of the commercial optical network has been used just once at the beginning of the experiment for alien lambdas provisioning. Later, we have relied just on the NETCONF/YANG management system of the two Infinera transponders. OPM data provided by the coherent receivers, shown in Table II, are downloaded every minute by an application based on the NETCONF protocol. The first set of 1440 OPM data records (1 record per minute during a 24 hour time period) was collected in the normal operating condition and labelled as the baseline (BSL) of our experiment. After full characterization of the system baseline, we have applied the following attack techniques.

1) In-band jamming attack: the intrusion signal is a CW low power signal whose frequency falls within the bandwidth of the signal under test.
2) Out-of-band jamming attack: the intrusion signal is a CW signal with a frequency outside the bandwidth of the signal under test.
3) Polarization modulation attack: the polarization modulator is activated and causes transmission errors as soon as the induced polarization variation is faster than the coherent receiver's polarization recovery algorithm (the intrusion signal is switched off).

A. *In-Band Jamming Attack*

The power spectrum of the OCh under test for the in-band attack is shown in Fig. 7b. The reader can notice that the jamming signal (the small peak on the right-hand side of the OCh spectrum) is slightly detuned with respect to the central frequency of the OCh under test as doing so makes the jamming particularly effective (i.e. a remarkable increase in BER-POST-FEC can be achieved by modest intrusion signal power).

We have implemented two in-band attack conditions, light (INBLGT) and strong (INBSTR), by setting the power of the intrusion signal to 10 and 7 dB below the power of the signal under test, respectively. In these conditions, the system produces Uncorrected Block Errors (UBE-FEC) ranging from a few blocks per minute to many thousands blocks per minute which we classify as light and strong attack conditions respectively. A full OPM data set with 1440 records has been logged for each of these two attack conditions.

B. *Out-of-Band Jamming Attack*

Fig. 7c shows the full DWDM power spectrum of the out-of-band attack. The intrusion signal frequency was 195.1 THz (1536.61 nm). In this type of attack, the intrusion signal is not included in the OCh plan of the network management system. When it is injected in the link, it represents an unforeseen input signal for the following optical amplifier and it creates a power reduction on the other operating OChs when the amplifier is

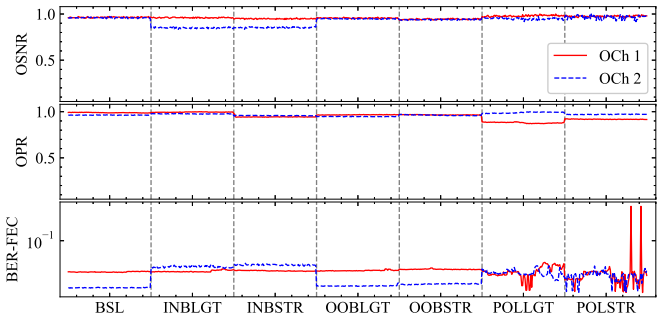


Fig. 8. A subset of OPM data collected during the considered security regimes: OSNR, OPR (both normalized) and BER-FEC.

configured in the constant power mode. This in turn produces a reduction of the OSNR that impairs all the working channels or degrades their performance. This feature distinguishes the out-of-band attack from the in-band one: all OChs passing through the link under attack are affected in the former, while just one OCh is harmed in the latter. On the other hand, in the out-of-band attack the power required for the intrusion signal is much higher than in the in-band attack: we have set the power of the intrusion signal to +3 and +8.7 dB with respect to the power of the OCh under test to get light (OOBLGT) and strong (OOBSTR) attack conditions respectively. A full OPM data set with 1440 records has been logged for each condition.

C. Polarization Modulation Attack

In the polarization attack, we have switched off the intrusion signal and activated the polarization state modulator. In this operating condition, all the optical parameters of the system are the same as the baseline condition, but the polarization modulation causes transmission errors when the coherent receiver's polarization recovery algorithm is unable to track the fast polarization changes [40]. Similarly to the out-of-band attack, the polarization attack affects all the coherent OChs traversing the link where polarization modulation is applied.

We have experimentally identified one of the resonant frequencies of the fiber squeezer at 136 kHz by monitoring the amplitude of the sinewave driving signal. Then, the sinewave amplitude was set to 0.4 and 1.6 V peak-to-peak resulting in light (POLLGT) and strong (POLSTR) attack, respectively. A full OPM data set with 1440 records has been logged for each condition.

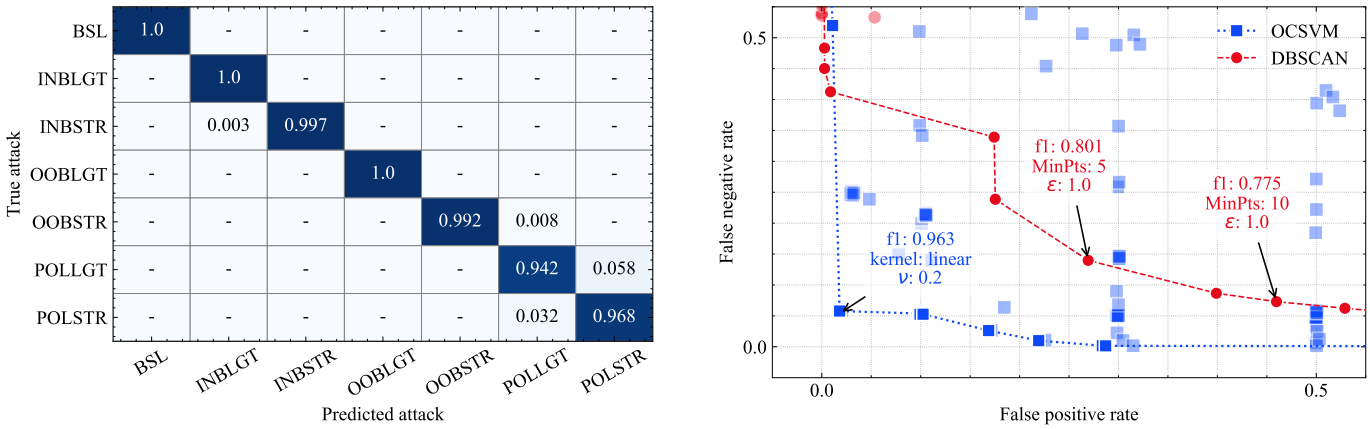
Fig. 8 provides an insight into general trends observed for representative OPM parameters, i.e., OSNR, OPR and BER-FEC collected during the considered security regimes. As can be seen in the figure, OSNR and OPR (whose values are normalized) show very little time variations over the different scenarios. BER-FEC shows a remarkable increase with respect to the baseline in in-band jamming attacks but remains almost unchanged in out-of-band jamming attacks, while it exhibits a noisy up-and-down behaviour in polarization modulation attacks. This weird behaviour prohibits the use of a threshold based approach for security diagnostics and makes identifying the root cause of the detected anomaly extremely difficult even for an optical system expert.

VII. Performance Analysis

The performance of the ML techniques described in Sec. III was evaluated using the dataset described in Sec. VI. In particular, ANN was chosen for SL, One-Class Support Vector Machine (OCSVM) for SSL and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) for UL, due to their state-of-the-art performance in many tasks [41]. The evaluation was performed using Python and the Scikit-learn implementation. The dataset contains 1440 samples from each connection under each attack condition, yielding a total of 20,160 samples. This is 7 times larger than in [8] and makes the largest experimental dataset related to optical network security reported to date. The collected samples were pre-processed by removing samples with missing data and applying z-score standardization.

For the ANN, several architectures were evaluated using k-fold cross validation technique, with the best one having 3 layers with [50, 100, 50] neurons. This architecture was then trained using a 50%, 25%, 25% training, validation and test sets for 1,000 epochs with Adam optimizer minimizing its categorical cross-entropy loss. For the OCSVM, three parameters were evaluated: kernel, which defines the kernel function adopted; ν, which defines an upper bound on the training errors; and g, which (when applicable) defines the coefficient of the kernel function. For the DBSCAN, two parameters were evaluated: $MinPts$, which defines the minimum number of samples within an area required to form a cluster; and ϵ, which defines the radius of the area of search. Both OCSVM and DBSCAN were evaluated using samples selected from the dataset with 10:1.5 normal-to-attack sampling ratio. To improve the reliability of the false positive and false negative rates reported for SSL and UL, attack detection was performed for 50 randomly sampled batches of NOC samples. For each of these NOC batches, we further sample 50 batches from each attack condition.

Fig. 9 presents the test results for the three evaluated models. Fig. 9a shows the confusion matrix of the test dataset after training the ANN. This matrix assesses the per-class accuracy of the ANN by showing the portion of samples from one class being correctly or wrongly predicted as another. The very high accuracy observed, with no false positives nor false negatives, translates into the maximum $f1$ score, i.e., 1. These results for the network scenario are compatible with the ones obtained for the single-connection scenario in [4]. Fig. 9b shows the false positive and false negative rates for OCSVM and DBSCAN. Each point corresponds to a particular configuration of the algorithm, while the line represents the Pareto frontier, i.e., the lowest false positive rate obtained for a particular false negative rate. Points outside the line are regarded as dominated points, i.e., they do not represent a good trade-off between the false positive and false negative rates. For OCSVM, many points are not a part of the Pareto frontier, while for DBSCAN most of the points fall on the highlighted curve. This is expected since the DBSCAN parameters clearly offer a trade-off between false positive and false negative rates. On the other



(a) Supervised learning (ANN)

(b) Semi-supervised (OCSVM) and unsupervised (DBSCAN) learning

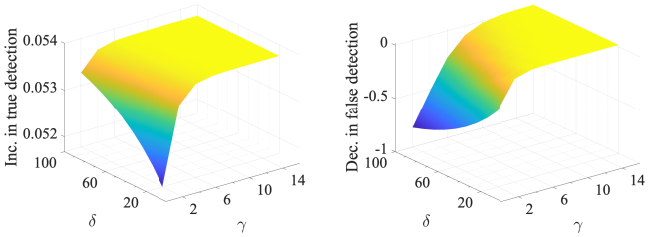
Fig. 9. Test results for the ML techniques tested. SL (a) is evaluated using the confusion matrix over the attack scenarios described in Sec. VI. SSL and UL (b) are evaluated by the obtained false positive and negative rates. Insets show the best configuration parameters for OCSVM and DBSCAN.

hand, OCSVM configurations result with more scattered points due to the different characteristics of the kernel functions. The OCSVM approach shows good accuracy, where the best configuration, i.e., the one with the highest $f1$ score of 0.963, achieves 1.7% false positives and 5.3% false negatives. On the other hand, DBSCAN has much lower accuracy, with the best trade-off between false positives and false negatives (i.e., the highest $f1$ score of 0.8) observed when $MinPts$ is set to 5, which results in a 26.9% false positive and a 13.9% false negative rate. Fig. 9b also shows a second point for DBSCAN where $MinPts$ is increased to 10, which prevents the algorithm from forming clusters with a fewer number of samples. This configuration reduces the false negatives (from 13.9% to 7.3%) at the expense of increasing the false positives (from 26.9% to 45.9%) but might represent a good trade-off in terms of security if network operators want to reduce the risk of attacks remaining undetected.

Considering the false positive and false negative rates achieved by each ML model, we focus on evaluating the benefits and drawbacks of the WAD proposed in Sec. III-A. In particular, we investigate the impact of the δ, γ and τ parameter values on the performance gains obtained by the WAD. As the ANN does not generate any false positives nor false negatives, we focus on overcoming the inaccuracies of SSL and UL.

Fig. 10 shows the benefits obtained by the WAD approach when applied to the false positive and false negative rates from the best OCSVM configuration, for varying δ and γ, with $\tau = \gamma/2$. As shown in Fig. 10a, the 5.3% false negative rate initially obtained by OCSVM is completely compensated for in most of the δ and γ configurations. By using higher values of γ, WAD averts degradation of the initially low false positive rate (i.e., 1.7%) obtained by OCSVM, as shown in Fig. 10b.

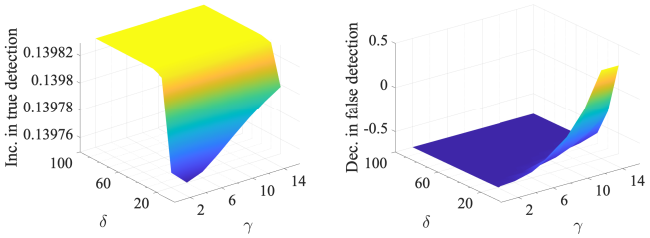
Fig. 11 shows the benefits of applying WAD to the best-performing configuration of DBSCAN ($F_P = 26.9\%$ and $F_N = 13.98\%$). The 13.98% false negative rate obtained by sample-based DBSCAN is compensated in all cases, with very small variations between configurations, as shown in Fig. 11a.



(a) Increase in true detection

(b) Decrease in false detection

Fig. 10. Benefits obtained by the Window-based Attack Detection (WAD) for different values of δ and γ, with τ=$\gamma/2$, considering the best OCSVM configuration.



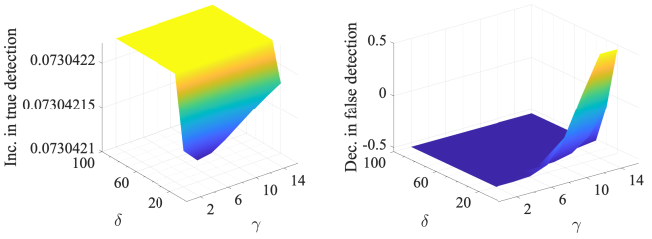
(a) Increase in true detection

(b) Decrease in false detection

Fig. 11. Benefits obtained by the WAD for different values of δ and γ, with τ=$\gamma/2$, considering the best DBSCAN configuration.

Moreover, as depicted in Fig. 11b, the 26.9% false positive rate can be transformed into zero falsely detected alarms by setting a low δ and a high γ. Therefore, the best configuration for the WAD in this scenario would be δ=20 and γ=16.

Fig. 12 shows the benefits obtained by the WAD approach when considering a more extreme case in which the DBSCAN algorithm is set to trade a low false negative rate (i.e., 7.3%) for a high false positive rate (i.e., 45.9%). This scenario is applicable when operators prefer to sacrifice false positive rate in favor of reducing the risk of attacks passing undetected. With similar configuration as in the previous case, WAD is able to maintain the good true positive properties of this DBSCAN



(a) Increase in true detection (b) Decrease in false detection

Fig. 12. Benefits obtained by the WAD for different configurations, with τ=$\gamma/2$, considering the DBSCAN with $MinPts$=10 and ϵ=1.0.

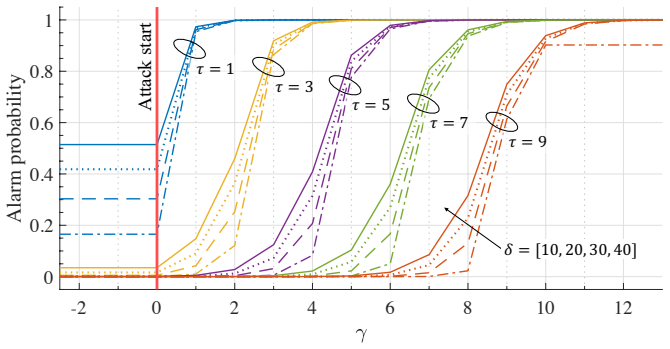


Fig. 13. Probability of raising an alarm over time samples γ for different values of attack detection window δ (shown in different line styles) and attack samples threshold τ (shown in different colors).

configuration while drastically reducing the false positives. These results demonstrate the effectiveness of the proposed window-based approach, and enable SSL and UL models to be reliably used during the optical network operation. Moreover, with a proper configuration of the WAD parameters, more reliable information is fed to the ALM for breach localization, improving the overall accuracy or security diagnostics.

Fig. 13 shows the probability of WAD raising an alarm over the number of samples collected from the start of an attack (γ). Assuming a monitoring cycle of 1 minute, the γ value can be translated into the number of minutes elapsed since the attack started, associated with the probability of raising an alarm. The values shown in the figure are obtained assuming the false positive and negative rates achieved by OCSVM. For $\gamma \leq 0$ no attack is present in the network, and the shown value refers to the false detection probability. The figure indicates that requiring fewer samples classified as attack (e.g., τ=1) increases the chances of detecting the attack in very early stages (i.e., the probability of detecting the attack with just the first sample exceeds 90%), but at the expense of a high false detection probability (from 16% when δ=10 up to 52% when δ=40). On the other hand, stipulating a large number of needed positive samples (e.g., τ=9) may impose an undesirable delay in detecting the attack. Balancing the τ values can achieve an adequate trade-off between the false detection probability and the time to raise an alarm upon an attack.

The two most promising configurations are τ=3 and τ=5, and their usage should be decided by the network operator depending on the overhead caused by false detection. If the overhead caused by false positives is not too high, τ=3 provides quicker detection at a very low false detection rate. For example, when τ=3 (δ=10), the attack detection probability upon receiving three attack samples equals 86%, and increases to 99% and 99.9% for the fourth and fifth attack sample, respectively. This is accompanied with a false detection probability of 0.063%. On the other hand, if the overhead caused by a false alarm is unacceptably high, the network operator might want to prevent false detection at the expense of a slightly longer time to detect attacks. This trade-off is obtained by setting τ=5, yielding zero false detection probability, and 78%, 97% and 99% of detection probability upon receiving the fifth, sixth an seventh attack sample, respectively. The figure also shows that the size of the observation window δ does not affect the performance of the algorithm substantially, except for very low or very high values of τ (e.g., τ=1 or τ=9).

VIII. Conclusions

This paper proposed an ML-based framework for cognitive and autonomous security diagnostics of physical-layer security in optical networks. The framework comprises modules for detection and identification of attacks that can leverage on supervised, semi-supervised and unsupervised learning to detect attacks and, when applicable, identify their type and intensity; as well as a module for connection- and/or link-wise localization of attacks, incorporated into the NMS. We propose a method for enhancing the performance of the ML approaches with a window-based approach, which was shown to compensate for the effect of false positives and false negatives.

The paper provided insights into practical implementation and integration aspects of the proposed framework and evaluated its performance for experimental data from an optical network testbed subjected to different attacks, demonstrating strong potential to perform autonomous, cognitive security diagnostics. Future works and open questions include a deeper investigation of the robustness of the monitoring platform to variations in the "normal" failure conditions (i.e. capability of correctly distinguish between attacks and failures).

Acknowledgment

The authors would like to thank Roberto Morro for his invaluable support with the data acquisition application, Lena Wosinska for fruitful discussions, and Infinera for providing the Groove G30 transponders.

References

[1] M. Furdek, C. Natalino, F. Lipp, D. Hock, N. Aerts, M. Schiano, A. Di Giglio, and L. Wosinska, "Demonstration of machine-learning-assisted security monitoring in optical networks," in *Proc. of ECOC*, Sept 2019.

[2] N. Skorin-Kapov, M. Furdek, S. Zsigmond, and L. Wosinska, "Physical-layer security in evolving optical networks," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 110–117, Aug 2016, DOI: 10.1109/MCOM.2016.7537185.

[3] T. Uematsu, H. Hirota, T. Kawano, T. Kiyokura, and T. Manabe, "Design of a temporary optical coupler using fiber bending for traffic monitoring," *IEEE Photonics J.*, vol. 9, no. 6, pp. 1–13, Dec 2017, DOI: 10.1109/JPHOT.2017.2762662.

[4] C. Natalino, M. Schiano, A. Di Giglio, L. Wosinska, and M. Furdek, "Experimental study of machine-learning-based detection and identification of physical-layer attacks in optical networks," *IEEE/OSA J. Lightwave Technol.*, vol. 37, no. 15, pp. 4173–4182, Aug 2019, DOI: 10.1109/JLT.2019.2923558.

[5] T. Tanaka, A. Hirano, S. Kobayashi, T. Oda, S. Kuwabara, A. Lord, P. Gunning, O. Gonzáles de Dios, V. Lopez, A. M. Lopez de Lerma, and A. Manzalini, "Autonomous network diagnosis from the carrier perspective [invited]," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 12, no. 1, pp. A9–A17, Jan 2020, DOI: 10.1364/JOCN.12.0000A9.

[6] D. Rafique, T. Szyrkowiec, H. Grießer, A. Autenrieth, and J.-P. Elbers, "Cognitive assurance architecture for optical network fault management," *IEEE/OSA J. Lightwave Technol.*, vol. 36, no. 7, pp. 1443–1450, Apr 2018, DOI: 10.1109/JLT.2017.2781540.

[7] L. Velasco, A. C. Piat, O. González, A. Lord, A. Napoli, P. Layec, D. Rafique, A. D'Errico, D. King, M. Ruiz, F. Cugini, and R. Casellas, "Monitoring and data analytics for optical networking: Benefits, architectures, and use cases," *IEEE Network*, pp. 1–9, 2019, DOI: 10.1109/MNET.2019.1800341.

[8] Y. Li, N. Hua, Y. Yu, Q. Luo, and X. Zheng, "Light source and trail recognition via optical spectrum feature analysis for optical network security," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 982–985, May 2018, DOI: 10.1109/LCOMM.2018.2801869.

[9] F. Paolucci, A. Sgambelluri, F. Cugini, and P. Castoldi, "Network telemetry streaming services in SDN-based disaggregated optical networks," *IEEE/OSA J. Lightwave Techn.*, vol. 36, no. 15, pp. 3142–3149, Aug 2018, DOI: 10.1109/JLT.2018.2795345.

[10] K. Christodoulopoulos, C. Delezoide, N. Sambo, A. Kretsis, I. Sartzetakis, A. Sgambelluri, N. Argyris, G. Kanakis, P. Giardina, G. Bernini, D. Roccato, A. Percelsi, R. Morro, H. Avramopoulos, P. Castoldi, P. Layec, and S. Bigo, "Toward efficient, reliable, and autonomous optical networks: the ORCHESTRA solution [invited]," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 11, no. 9, pp. C10–C24, Sep 2019, DOI: 10.1364/JOCN.11.000C10.

[11] D. Rafique and L. Velasco, "Machine learning for network automation: Overview, architecture, and applications [invited tutorial]," *IEEE/OSA J. Optical Commun. Network.*, vol. 10, no. 10, pp. D126–D143, Oct 2018, DOI: 10.1364/JOCN.10.00D126.

[12] J. Mata, I. de Miguel, R. J. Durán, N. Merayo, S. K. Singh, A. Jukan, and M. Chamania, "Artificial intelligence (AI) methods in optical networks: A comprehensive survey," *Optical Switching and Networking*, vol. 28, pp. 43–57, Apr 2018, DOI: 10.1016/j.osn.2017.12.006.

[13] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An optical communication's perspective on machine learning and its applications," *IEEE/OSA J. Lightwave Techn.*, vol. 37, no. 2, pp. 493–516, Jan 2019, DOI: 10.1109/JLT.2019.2897313.

[14] F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1383–1408, Secondquarter 2019, DOI: 10.1109/COMST.2018.2880039.

[15] G. Choudhury, D. Lynch, G. Thakur, and S. Tse, "Two use cases of machine learning for SDN-enabled IP/optical networks: traffic matrix prediction and optical path performance prediction [invited]," *IEEE/OSA J. Optical Commun. Network.*, vol. 10, no. 10, pp. D52–D62, Oct 2018, DOI: 10.1364/JOCN.10.000D52.

[16] I. Sartzetakis, K. K. Christodoulopoulos, and E. M. Varvarigos, "Accurate quality of transmission estimation with machine learning," *IEEE/OSA J. Optical Commun. Network.*, vol. 11, no. 3, pp. 140–150, March 2019, DOI: 10.1364/JOCN.11.000140.

[17] T. Panayiotou, G. Savva, B. Shariati, I. Tomkos, and G. Ellinas, "Machine learning for QoT estimation of unseen optical network states," in *Proc. OFC*, March 2019, p. Tu2E.2.

[18] Z. Wang, M. Zhang, D. Wang, C. Song, M. Liu, J. Li, L. Lou, and Z. Liu, "Failure prediction using machine learning and time series in optical network," *Opt. Express*, vol. 25, no. 16, pp. 18 553–18 656, Aug 2017, DOI: 10.1364/OE.25.018553.

[19] S. Shahkarami, F. Musumeci, F. Cugini, and M. Tornatore, "Machine-learning-based soft-failure detection and identification in optical networks," in *Proc. of OFC*, 2018, p. M3A.5.

[20] B. Shariati, M. Ruiz, J. Comellas, and L. Velasco, "Learning from the optical spectrum: Failure detection and identification," *IEEE/OSA J. Lightwave Techn.*, vol. 37, no. 2, pp. 433–440, Jan 2019, DOI: 10.1109/JLT.2018.2859199.

[21] X. Chen, B. Li, R. Proietti, Z. Zhu, and S. J. B. Yoo, "Self-taught anomaly detection with hybrid unsupervised/supervised machine learning in optical networks," *IEEE/OSA J. Lightwave Techn.*, vol. 37, no. 7, pp. 1742–1749, April 2019, DOI: 10.1109/JLT.2019.2902487.

[22] F. Musumeci, C. Rottondi, G. Corani, S. Shahkarami, F. Cugini, and M. Tornatore, "A tutorial on machine learning for failure management in optical networks," *IEEE/OSA J. Lightwave Techn.*, vol. 37, no. 16, pp. 4125–4139, Aug 2019, DOI: 10.1109/JLT.2019.2922586.

[23] M. P. Fok, Z. Wang, Y. Deng, and P. R. Prucnal, "Optical layer security in fiber-optic networks," *IEEE Inf. Foren. Sec.*, vol. 6, no. 3, pp. 725–736, April 2011, DOI: 10.1109/TIFS.2011.2141990.

[24] N. Skorin-Kapov, J. Chen, and L. Wosinska, "A new approach to optical networks security: Attack-aware routing and wavelength assignment," *IEEE Trans. Netw.*, vol. 18, no. 3, pp. 750–760, June 2010, DOI: 10.1109/TNET.2009.2031555.

[25] N. Skorin-Kapov, M. Furdek, R. Aparicio-Pardo, and P. Pavón-Mariño, "Wavelength assignment for reducing in-band crosstalk attack propagation in optical networks: ILP formulations and heuristic algorithms," *European J. Oper. Res.*, vol. 222, no. 3, pp. 418–429, Nov 2012, DOI: 10.1016/j.ejor.2012.05.022.

[26] K. Manousakis, P. Kollios, and G. Ellinas, "Multi-period attack-aware optical network planning under demand uncertainty," in *Optical Fiber and Wireless Communications*, R. Roka, Ed., June 2017, DOI: 10.5772/intechopen.68491.

[27] J. Zhu, B. Zhao, and Z. Zhu, "Leveraging game theory to achieve efficient attack-aware service provisioning in EONs," *IEEE/OSA J. Lightwave Techn.*, vol. 35, no. 10, pp. 1785–1796, May 2017, DOI: 10.1109/JLT.2017.2656892.

[28] A. Saltykov, S. Glagolev, J. B. Jensen, and I. T. Monroy, "Security attacks in optical access networks - simultaneous detection and localization," in *Proc. of IEEE Photonic Society 24th Annual Meeting*, Oct. 2011, pp. 935–936, DOI: 10.1109/PHO.2011.6110867.

[29] C. Mas, I. Tomkos, and O. K. Tonguz, "Failure location algorithm for transparent optical networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 8, pp. 1508–1519, Aug 2005, DOI: 10.1109/JSAC.2005.852182.

[30] Tao Wu and A. K. Somani, "Cross-talk attack monitoring and localization in all-optical networks," *IEEE/ACM Trans. Netw.*, vol. 13, no. 6, pp. 1390–1401, Dec 2005, DOI: 10.1109/TNET.2005.860103.

[31] Z. Dong, F. N. Khan, Q. Sui, K. Zhong, C. Lu, and A. P. T. Lau, "Optical performance monitoring: A review of current and future technologies," *IEEE/OSA J. Lightwave Technol.*, vol. 34, no. 2, pp. 525–543, Jan 2016, DOI: 10.1109/JLT.2015.2480798.

[32] F. Pederzolli, M. Furdek, D. Siracusa, and L. Wosinska, "Towards secure optical networks: A framework to aid localization of harmful connections," in *Proc. of OFC*, March 2018, p. Th2A.42.

[33] M. Furdek, V. W. S. Chan, C. Natalino, and L. Wosinska, "Network-wide localization of optical-layer attacks," in *Proc. of ONDM*, Athens, Greece, May 2019, pp. 310–322, DOI: 10.1007/978-3-030-38085-4_27.

[34] Y. Li, N. Hua, Y. Song, S. Li, and X. Zheng, "Fast lightpath hopping enabled by time synchronization for optical network security," *IEEE Commun. Lett.*, vol. 20, no. 1, pp. 101–104, Jan 2016, DOI: 10.1109/LCOMM.2015.2497703.

[35] M. Furdek, N. Skorin-Kapov, and L. Wosinska, "Attack-aware dedicated path protection in optical networks," *IEEE/OSA J. Lightwave Techn.*, vol. 34, no. 4, pp. 1050–1061, Feb 2016, DOI: 10.1109/JLT.2015.2509161.

[36] E. Hugues-Salas, F. Ntavou, D. Gkounis, G. T. Kanellos, R. Nejabati, and D. Simeonidou, "Monitoring and physical-layer attack mitigation in SDN-controlled quantum key distribution networks," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 11, no. 2, pp. A209–A218, Feb 2019, DOI: 10.1364/JOCN.11.00A209.

[37] M. Bensalem, S. Kumar Singh, and A. Jukan, "On detecting and preventing jamming attacks with machine learning in optical networks," *arXiv:1902.07537 [cs.NI]*, Jun 2019.

[38] M. Furdek, C. Natalino, M. Schiano, and A. Di Giglio, "Experiment-based detection of service disruption attacks in optical networks using data analytics and unsupervised learning," in *Metro and Data Center Optical Networks and Short-Reach Links II*, vol. 10946, San Francisco, CA, USA, 2019, DOI: 10.1117/12.2509613.

[39] Infosim. (2018, May) StableNet – unified network and services management. [Online]. Available: https://www.infosim.net/stablenet/

[40] P. M. Krummrich, D. Ronnenberg, W. Schairer, D. Wienold, F. Jenau, and M. Herrmann, "Demanding response time requirements on coherent receivers due to fast polarization rotations caused by lightning events," *Opt. Express*, vol. 24, no. 11, pp. 12 442–12 457, May 2016, DOI: 10.1364/OE.24.012442.

[41] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?" *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3133–3181, Jan. 2014.