Cyber-physical event reasoning for distributed energy resources

Nils Müller[a,*], Kaibin Bao[b], Kai Heussen[a]

[a] Wind and Energy Systems Department, Technical University of Denmark, Building 330, Risø campus, 4000 Roskilde, Denmark
[b] Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Building 445, Campus North, 76344 Eggenstein-Leopoldshafen, Germany

Abstract

The widespread adoption of internet-connected and remotely controllable solar plants and energy storages renders coordinated cyber-physical attacks against distributed energy resources (DERs) an emerging risk for power systems. Effective incident response can be facilitated by online DER monitoring providing real-time information on event root causes and physical impacts. Such online event identification is challenged by the lack of historical attack observations, and emergence of new attack strategies. The **Cy**ber-**Ph**ysical **E**vent **R**easoning **S**ystem CyPhERS provides real-time information on both known and unknown attack types in form of informative and interpretable event signatures, without need to be trained on historical attack samples. To date, CyPhERS has only been demonstrated on a laboratory water distribution testbed of limited complexity, considering human evaluation of event signatures. This work methodologically adapts CyPhERS to specificities of DER operation such as weather and consumer-induced volatility, and introduces an automated signature evaluation system. The feasibility of applying CyPhERS for automated DER monitoring is investigated on a dataset recorded from a real photovoltaic-battery system targeted by several cyber and cyber-physical attack types. The results demonstrate that the proposed methodological adaptations and signature evaluation system enable the application of CyPhERS for automated online identification of different attack types targeting DERs, while greatly reducing the false positive rate.

Keywords: Distributed energy resources, PV-battery systems, Attack detection, Cyber-physical monitoring, Machine learning

1. Introduction

The transformation towards widespread use of sustainable energy sources is driven by decentralization and electrification. Both the replacement of centralized fossil power plants with renewable generation, as well as the electrification of the mobility and heating sectors are boosting the deployment of distributed energy resources (DERs) such as solar plants, electric vehicles, battery storages and heat pumps. The large-scale adoption of DERs provides benefits beyond decarbonizing energy consumption, including lower transmission costs and improved grid stability through provision of ancillary services [1]. Harnessing this potential requires integration with information and communication technology (ICT) for continuous coordination and management of numerous geographically distributed devices. However, the associated connection to public networks and remote control capability, combined with often low security standards [2], render DERs promising targets for cyber criminals. Incidents such as the Mirai botnet attack have demonstrated that a fleet of internet of things (IoT) devices can be simultaneously seized [3]. Malicious control of multiple DERs can provoke grid instability by switching the devices simultaneously on or off, rendering coordinated attacks on DERs a serious threat for power system operation [4]. In this context, the increasing number of attacks on critical infrastructure underlines the need to support the large-scale deployment of DERs with adequate security mechanisms [5, 6].

Attack detection is among the most frequently suggested security measures for DERs [5, 7]. Once an attack is detected and identified, affected systems and network zones can be isolated, and incident response mechanisms activated. In the light of cyber-physical attacks, timely and appropriate counteractions require real-time information on both root causes and physical impact. Most existing detection concepts exclusively monitor either cyber network traffic or physical process data [7]. While cyber network attack detection potentially allows to distinguish several attack types, physical impacts are not identified. In contrast, physical attack detection can determine the attack impact, but not the underlying attack vector. Consequently, some works propose the combined evaluation of operational technology (OT) network traffic and process data [7], and demonstrate the superior performance of such cyber-physical attack identification concepts applying supervised machine learning (ML) [8]. However, due to the dependency on historical samples of typically rarely occurring attacks, supervised methods lack practical relevance [9, 10]. In [11], the authors introduce CyPhERS, a **Cy**ber-**Ph**ysical **E**vent **R**easoning **S**ystem which exploits advantages of cyber-physical monitoring while being independent of historical attack observations. So far, CyPhERS has only been demonstrated on a laboratory water distribution testbed exhibiting simple repeating process patterns. Moreover, the demonstrated version of the concept requires active involvement of human operators. In the context of DER monitoring, the problem is more complicated due to the pronounced volatility and randomness resulting from dependency on weather and consumer behavior [10]. Moreover, especially for small scale DERs, active operator participation is impractical. Thus, this work addresses the following research question: *How can real-time information about cyber(-physical) attacks against DERs such as occurrence, type, victim devices, attacker location, and physical impact be provided in an automated fashion, while being independent of historical attack observations?*

For this purpose, the present study methodologically adapts CyPhERS to the specificities of DER operation, and introduces

*Corresponding author. E-mail address: nilmu@dtu.dk (N. Müller).

an automated signature evaluation system. Key features of the adaptation comprise switching from deterministic to probabilistic models and detection rules as well as considering and monitoring functional, behavioral and abstracting physical target features of a DER. The effectiveness of the adaptations and the automated signature evaluation system is demonstrated on a dataset derived from a real photovoltaic (PV)-battery system targeted by various cyber(-physical) attack types, and is supported by a quantitative performance comparison to the original version of CyPhERS.

1.1. Related work

An exhaustive review of literature related to CyPhERS, as well as a conceptual comparison and performance benchmarking with other event identification concepts is provided in [11]. This section specifically reviews works on attack detection and identification methods for DERs and other power system applications. These can be broadly divided into methods monitoring the cyber network, physical process or both.

1.1.1. Physical attack detection and identification

Many works propose attack detection applying physics-based models [12–16]. The models are used to emulate a DER under normal condition. By evaluating the residual between the model and actual measurements against a threshold, attacks can be detected [17]. An advantage of this approach is the independency from attack samples [9]. However, accurate modeling might be challenging for DERs with complex architectures (e.g., hybrid power plants) leading to imprecise detection. Moreover, the restriction to a binary detection problem (*normal* vs. *abnormal operation*) omits insights on root causes and physical impacts. Data-driven methods constitute another widely considered approach for physical attack detection and identification [7]. One argument is generalizability as process representations are automatically learned from data, avoiding expensive manual model development. The majority of works considers supervised approaches such as binary [18, 19] and multi-class classification [10, 20–23]. The explicit learning from attack samples, on the one hand, allows to detect and differentiate various cyber-physical attack types based on their physical impact. On the other hand, it renders supervised methods impractical due to the natural scarcity of such data. Other works apply regression or autoencoder models to learn the normal behavior of a DER, and detect attacks by comparing the model with the actual measurements [24–26]. Similar to the approaches applying physics-based models, the restriction to a binary detection problem makes them of limited use for incident response.

1.1.2. Cyber network attack detection and identification

Among the classical approaches for monitoring DER network data are signature-based intrusion detection systems applying tools such as Snort [27]. These can detect and differentiate attacks in case of known attack signatures. Related but newer approaches include supervised detection of attack patterns, for example, firmware modifications in inverter-based microgrids [28]. Neither the traditional signature-based nor the newer supervised ML-based methods can detect new attack strategies. Furthermore, they do not provide any information about the physical impact of an attack on the operation of a DER. Another approach is the detection of anomalies in network traffic, known as behavior-based intrusion detection [7]. In the recent years, an increasing focus is on ML-based normal behavior reference models, which are compared to actual traffic, allowing to detect anomalies [29–31]. Although such approaches can potentially detect both known and unknown types of attacks, they do not provide any information other than the occurrence of abnormal network behavior.

1.1.3. Cyber-physical attack detection and identification

As concepts which exclusively monitor either a DER's cyber or physical domain neither can identify both the attack root cause and physical impact nor accurately differentiate between cyber attacks, cyber-physical attacks, network failures, and process faults, many works suggest investigation of cyber-physical detection [7, 8, 32]. Nevertheless, literature on the combined evaluation of process and network data of a DER or other power system applications is rare. The authors of [33] propose joint evaluation of synchrophasor measurements and properties of network traffic applying a multi-class decision tree classifier. In [34], unsupervised anomaly detection is applied to both network traffic and physical process features of a DER. A comparison of cyber, physical and cyber-physical detection in power systems is conducted in [35] by applying both supervised and unsupervised methods for the binary detection problem (*normal* vs. *abnormal operation*). The listed works all indicate that the joint monitoring of cyber and physical DER data improves detection performance. However, none of them combines root cause and physical impact identification with independence of historical attack samples. The authors in [11] propose the cyber-physical event reasoning system CyPhERS (see Fig. 1) to close this gap. CyPhERS utilizes a two-stage process to deduce event information, including the occurrence, type, location, and physical impact, from joint processing of network traffic and physical process data in real-time. The first stage generates informative event signatures for both unknown and known types of cyber attacks and physical faults. This is achieved through a combination of several methods including cyber-physical data fusion, unsupervised multivariate time series anomaly detection, and anomaly type differentiation. In the second stage, the event signatures are evaluated either through automated matching with a database of known event signatures or through manual interpretation by the operator. While the authors claim that the evaluation of event signatures can be automated, only manual interpretation is realized to date. Moreover, the concept demonstration is conducted on a simple laboratory water distribution system. Thus, applicability for DER monitoring first needs to be demonstrated.

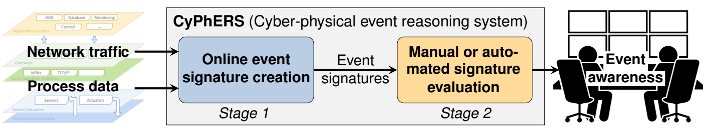


Figure 1: Schematic representation of CyPhERS adapted from [11].

1.2. Contribution and paper structure

The main contributions of this work are as follows:

- Methodological adaptation of CyPhERS to the operation of DERs, including switching to probabilistic models and detection rules, as well as monitoring of functional, behavioral, and abstracting target features.

- Introduction and realization of an automated event signature evaluation system in CyPhERS' Stage 2.

- Feasibility demonstration of applying the adapted version of CyPhERS for automated online identification of cyber(-physical) attacks targeting DERs on data of a real PV-battery system, including a quantitative performance comparison to the original version.

The remainder of the paper is structured as follows: In Section 2, CyPhERS is conceptually summarized. Section 3 presents the real PV-battery system, attack scenarios and recorded dataset which serve as demonstration case. The methodology of CyPhERS and its adaptation to DER monitoring is detailed in Section 4 together with the implementation for the considered PV-battery system case. In Section 5, results of applying the adapted version of CyPhERS to the demonstration case are presented, including a performance comparison to the original version. Finally, key findings of the demonstration are discussed in Section 6, followed by a conclusion in Section 7.

2. Introduction of the CyPhERS concept

This section provides a summary of the detailed conceptual introduction of CyPhERS included in [11]. The concept of the online event signature creation (Stage 1) is summarized in Section 2.1. Thereafter, Section 2.2 provides a conceptual overview of the signature evaluation (Stage 2). Methodological details of Stage 1 and 2 follow in Section 4.

2.1. *Online event signature creation (Stage 1)*

CyPhERS' Stage 1 (see Fig. 2) combines a range of concepts to produce informative and human-readable event signatures for known and unknown types of attacks and failures in an online fashion. The signatures encompass information including event occurrence, type, location, and physical impact. The applied concepts are introduced in the following.

2.1.1. *Fusion of cyber and physical information*

A key feature of Stage 1 is the joint monitoring and evaluation of physical process and cyber network data (see Fig. 2). The intention is to describe possible interactions between physical and network processes during detected events by means of the generated event signatures to facilitate the differentiation of cyber attacks, cyber-physical attacks, and physical failures in the subsequent signature evaluation (Stage 2).

2.1.2. *Feature-level monitoring*

The second concept is the individual monitoring and evaluation of multiple system variables of a DER and the representation of their potentially abnormal behavior in the event signatures. These cover both variables of multiple process or network components of a DER and multiple variables of the same component, as illustrated in Fig. 2. While the former allows to indicate affected DER components in a generated event signature, the latter further specifies abnormal behavior of the concerned device. The monitored variables are derived from sensor readings and OT network traffic, and in the following denoted *target features*, where $\mathcal{I}$ and $\mathcal{J}$ represent the physical and network feature subset, respectively. For a target feature c, its time series is given as $X_c = \{x_1^c, x_2^c, ..., x_N^c \mid x_i^c \in \mathbb{R}\ \forall i\}$. The extraction of target features and the related proposed methodological adaptations for DER monitoring are further detailed in Section 4.1.

2.1.3. *Unsupervised time series anomaly detection using covariates*

The third conceptual element of Stage 1 is the utilization of covariate-based unsupervised time series anomaly detection for monitoring the set of physical and network target features within the signature extraction system (see Fig. 2). First, a normal behavior reference model is derived for each target feature. Their predictions are then compared to actual observations to detect abnormal behavior of individual target features. The key argument for applying unsupervised anomaly detection is the independence of historical event observations, which allows to indicate occurrence of both known and unknown event types in the event signatures. The benefit of monitoring target features as time series is the detection of deviations from normal behavior which are only abnormal in a

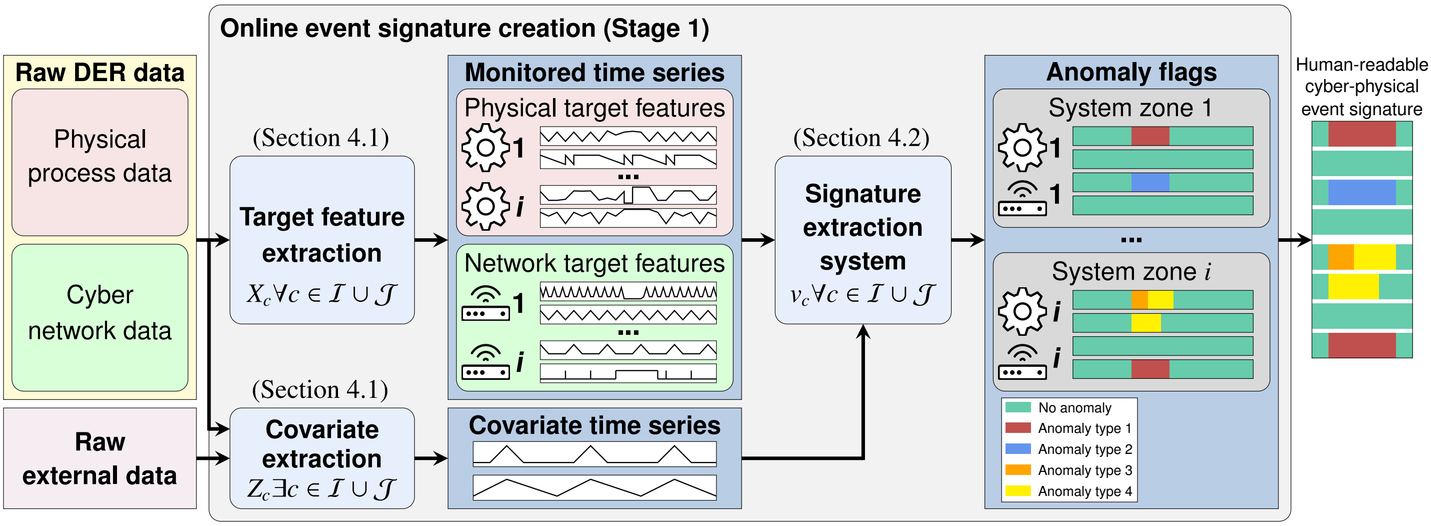


Figure 2: Conceptual overview of CyPhERS' online event signature creation (Stage 1) based on [11].

specific temporal context (local anomalies) [36]. Additionally, covariates are used to provide the normal behavior reference models with further DER internal or external information, allowing detection of situational anomalies which are only abnormal in the context of the provided covariates (e.g., detecting abnormal PV feed in context of irradiation). A covariate time series associated with a target feature c is formally denoted $Z_c = \{z_1^c, z_2^c, ..., z_N^c \mid z_i^c \in \mathbb{R}\ \forall i\}$ in the following. A detailed description of the applied anomaly detection methodology, including the proposed adaptation for DER monitoring, is provided in Section 4.2.

2.1.4. Differentiation of anomaly types

The fourth key feature of Stage 1 pertains to the differentiation of multiple anomaly types and their representation in the event signatures. Once an anomaly is detected for a target feature c, it is further classified using characteristics such as the direction of the deviation (e.g., abnormally *low* PV feed). The anomaly types are represented within the generated event signatures by different colors, enabling simple recognition and differentiation by humans (see Fig. 2). This distinction of anomaly types facilitates identification of event root causes and physical impacts in the subsequent signature evaluation (Stage 2). The series of anomaly flags produced by the signature extraction system for a target feature c is represented as $v_c = \{v_1^c, v_2^c, ..., v_N^c \mid v_i^c \in \mathbb{Z}\ \forall i\}$. The methodology of the anomaly type differentiation is further specified in Section 4.2.

2.1.5. Anomaly flags organization as readable event signatures

The fifth conceptual feature of Stage 1 is the joint visualization of the detection results of all target features as event signatures which can be easily interpreted by humans to identify included information. As previously described, Stage 1 takes multiple domains, system variables, and anomaly types into account to provide dense event information in form of anomaly flag series of a set of target features. To ease readability and information extraction, these flag series are re-organized by grouping them for each system zone of a DER (see Fig. 2). A system zone is defined as a collection of process and network components that are functionally linked (e.g., a battery stack and the associated smart battery inverter). Their logical relation facilitates associating anomaly flags of different target features. Consequently, Stage 1 of CyPhERS generates event signatures that are both rich in information and easily interpretable by humans.

2.2. Signature evaluation (Stage 2)

Fig. 3 illustrates the concept of CyPhERS' Stage 2. Stage 2 is concerned with the evaluation of event signatures provided by Stage 1, which can be performed by human operators or automated evaluation systems. The signatures are distinguishable and specific to event types. For known attack or fault vectors, they can be pre-defined and stored in a database. Once Stage 1 detects an event, the provided signature can be compared to the database. If a match is found, information such as the type of event, the affected components, the attacker location, and the physical impact can be retrieved to form a hypothesis about the event. Matching of signatures can be done by visual comparison or an automated evaluation system. One automation approach is the transformation of a pre-defined signature into a set of rules in the following form: ***flagging of*** *anomaly type 1* ***in*** *target feature X* ***and*** *type 2* ***in*** *target feature Y* ***indicates*** *device A* ***being targeted by*** *attack type B* ***causing*** *physical impact C*.

Importantly, signatures can also be defined for unknown event types based on partial knowledge, for example, ***flagging of*** *anomaly type 1* ***in*** *target feature X* ***indicates*** *device A failure [type unknown]* ***causing*** *physical impact B*. Consequently, even without the expert knowledge required to pre-define and recognize signatures of specific event types, or in case of new attack vectors, basic event information such as the occurrence, affected devices, and physical impact can be provided.

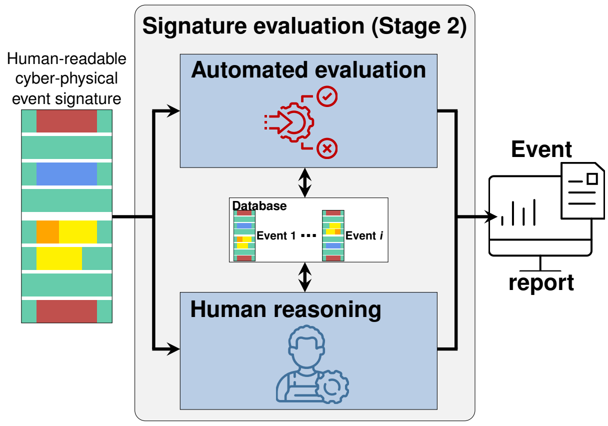


Figure 3: Conceptual overview of CyPhERS' signature evaluation (Stage 2) adapted from [11].

3. Demonstration case description

This section introduces the demonstration case considered for evaluating the effectiveness of the proposed methodological adaptations and automated signature evaluation system for applying CyPhERS for automated online DER monitoring. The demonstration case is based on cyber attack experiments on a real DER system, allowing to evaluate CyPhERS under realistic conditions and without assumptions and simplifications associated with simulation-based studies for the first time. The real PV-battery system which builds the foundation of the demonstration case is detailed in Section 3.1. Thereafter, Section 3.2 describes the considered attack scenarios. Finally, Section 3.3 details the realization of the attack experiments on the real PV-battery system, and the resulting dataset on which the CyPhERS demonstration in Section 5 is based.

3.1. PV-battery system

The cyber-physical structure of the considered real PV-battery system is schematically depicted in Fig. 4. The system is located at the Karlsruhe Institute of Technology (KIT), Germany. The system comprises four PV inverters, each with four dedicated solar panel strings (PV1-4), four battery stacks with associated battery inverters (BAT1-4), four energy meters (M1-4), a data manager (DM), and a data server (DS). The communication is based on modbus (MB), transmission control protocol (TCP), address resolution protocol (ARP), and user datagram protocol (UDP). Each battery stack has a capacity of 10.24 kWh and a maximum dis-/charge power of 5 kW. The connected solar panel strings of each inverter have a peak power between 15.50 kW_p and 16.74 kW_p. While PV1-4 are connected to all phases (L1-L3), BAT1-4 are linked to individual lines (see Fig. 4). The three phases are measured individually by M1-3. In addition, M4 measures all three

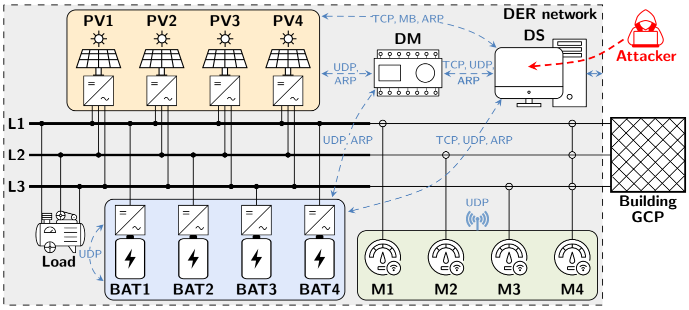


Figure 4: Illustration of the real PV-battery system considered for demonstrating the application of CyPhERS for DER monitoring.

phases and thus provides measurement redundancy. The PV and battery inverters are connected to the same grid connection point (GCP) as the building load. The load is characterized by typical office building patterns (higher load during working hours, lower on weekends). An additional characteristic are periodic load peaks due to activation of an air compressor. The control objective of the system is to minimize active power exchange with the grid. Consequently, batteries are charged if PV production exceeds the load, and discharged in the opposite case, provided an appropriate state of charge (SOC). As BAT1-4 are connected to individual phases, power flows P^{L1}-P^{L3} are controlled separately by a dedicated battery. An exception is P^{L1}, which is connected to BAT1 and BAT4. In case that batteries reach its maximum dis-/charging power, the offset on the respective phase is compensated by the other batteries on their phase, which is coordinated by communication among BAT1-4. Fig. 5 illustrates the operation of the batteries on the example of BAT3 for a representative day. At ~09:30, P^{L3} changes from grid import to export due to the increasing PV feed. Consequently, BAT3 starts charging to minimize the grid exchange. Between ~11:00 and ~17:00, the maximum charging power and SOC of BAT3 are reached, resulting in a deviation of P^{L3} from zero. At ~17:00, P^{L3} changes back from export to import due to the decreasing PV feed. As a result, BAT3 discharges to minimize the grid exchange until it is fully depleted at ~20:00.

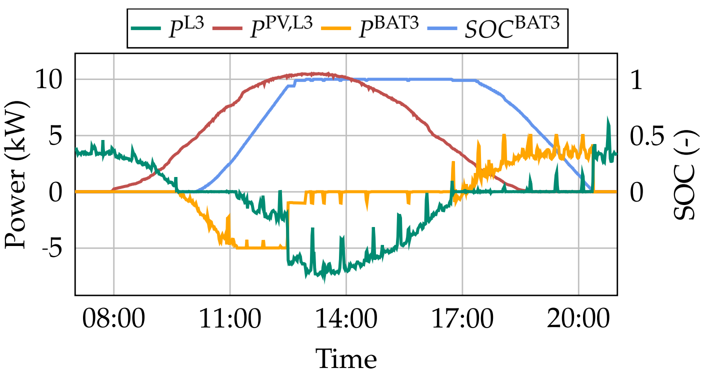


Figure 5: Illustration of the operation of the batteries on the example of BAT3.

The battery controllers receive the required measurements of P^{L1}-P^{L3} through subscription to the UDP multicast of the respective energy meter (M1, M2 or M3). The DM collects measurements from solar panels and batteries such as panel and cell temperatures. The DS hosts a custom data visualization software for users and builds the interface to the external network.

3.2. Attack scenarios

For the demonstration case, a threat scenario is assumed in which the attacker gained virtual access to the local DER network by hijacking the DS (see Fig. 4). From there she/he launches several cyber and cyber-physical attacks targeting different devices of the PV-battery system. The considered attack types are among the most relevant ones for DERs [10, 37, 38]. The cyber attacks comprise synchronize (SYN) scans and hypertext transfer protocol secure (HTTPS) requests, falling under the category of reconnaissance activities, as well as ARP spoofing used for eavesdropping, which belongs to data collection activities [39]. Among the cyber-physical attacks are false data injection attacks (FDIAs), false command injection attacks (FCIAs), and replay attacks, which all intent to alter the power output of the PV-battery system, as illustrated in Fig 6. In case of the FDIAs, false active power readings are injected in the name of the respective meter, causing an abrupt dis-/charging process of the batteries. The FCIAs comprise shutdown of either PV or battery inverters. For the replay attacks, the attacker repeats valid active power readings of the energy meters, which multiplies the control error and thus results in oscillation of the batteries. Given a simultaneous manipulation of multiple DERs, the considered cyber-physical attacks could provoke a sudden static (FDIA and FCIA) or dynamic (replay attack) load altering in the power system, which can cause congestion as well as voltage and frequency stability issues [40]. Attacks with the intent to create power quality disturbances on a millisecond or waveform scale are outside the scope of this work as those cannot be captured with the given data resolution of the considered real PV-battery system (see Section 3.3).

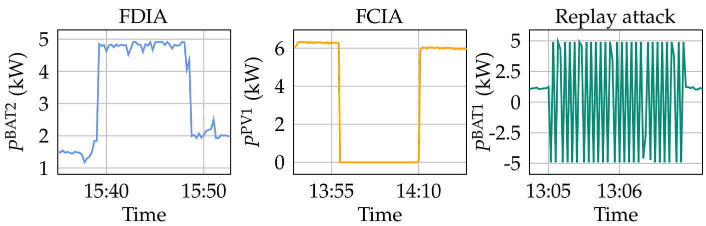


Figure 6: Physical impact of the considered cyber-physical attacks.

3.3. Attack experiments and dataset

The experiments have been conducted in October 2022. After recording the system under normal operation for approximately two weeks, 15 attacks were launched within one day between 10:00 and 17:00 local time (see Table 1).

The data was recorded using port mirroring in form of a passive packet capture of the local network. Both physical process and network traffic features are extracted from the resulting pcap file. The set of considered raw features is listed in Table 2. The physical data have a resolution between one second and one minute, depending on the respective feature. Network data on average comprise 7539 packets per minute.

4. Methodological adaptation and implementation of CyPhERS

This section details the methodology of CyPhERS with a particular focus on the proposed adaptation for DER monitoring and

Table 1: Schedule of the attack experiments.

No.	Attack type	Victim	Start	End
1	FDIA	M3	10:01	10:17
2	ARP spoof	PV3/DM	10:31	10:48
3	HTTPS request	BAT1	11:06	11:06
4	SYN Scan	PV3	11:22	11:34
5	FCIA	PV2	11:47	12:01
6	FDIA	M1	12:14	12:32
7	HTTPS request	DM	12:47	12:47
8	Replay attack	M1	13:05	13:07
9	FCIA	BAT3	13:26	13:39
10	FCIA	PV1	13:56	14:09
11	ARP spoof	PV4/DM	14:30	14:44
12	FCIA	BAT4	15:00	15:19
13	FDIA	M2	15:39	15:48
14	SYN Scan	PV4	16:04	16:08
15	Replay attack	M2	16:20	16:23

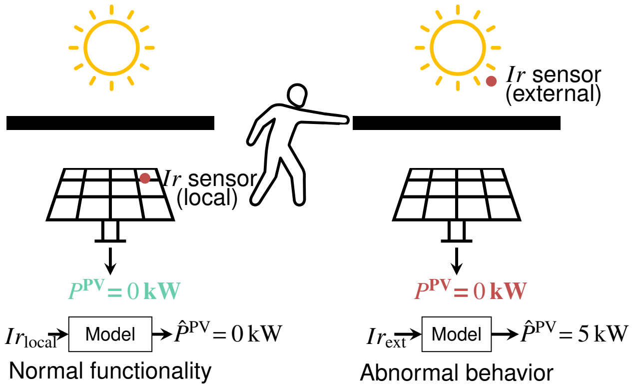


Figure 7: Comparison of functionality- and behavior-describing target features on the example of a physical "attack" where PV panels are covered.

automated signature evaluation system. First, the online event signature creation (Stage 1) is addressed, which includes a description of target feature and covariate extraction (Section 4.1), as well as the signature extraction system (Section 4.2). In this context, the central adaptations are presented which comprise 1) the switch from deterministic to probabilistic models and detection rules, 2) the consideration and differentiation of functional and behavioral physical target features, and 3) the use of abstracting physical target features. Thereafter, the signature evaluation (Stage 2) is explained in Section 4.3, which involves introduction of the proposed automated signature evaluation system. Along the methodological description of Stage 1 and 2, details on the implementation to the PV-battery system demonstration case are provided.

4.1. Target feature and covariate extraction

In this section, the extraction of target features and covariates from raw data of a DER is presented. Physical and network target features are addressed sequentially in the Sections 4.1.1 and 4.1.2.

4.1.1. Physical target features

According to [11], physical target features are monitored to identify both true physical events and manipulations of process-relevant data. The former requires features which represent the operation of all physical components of the system in question in order to localize the affected ones, and derive the physical impact on them. The latter necessitates monitoring of sensor readings used for process control. Therefore, this work considers physical target features which 1) represent the physical operation of PV1-4, BAT1-4, and M1-3, and 2) monitor the multicasted active power readings required for controlling the batteries. A specificity of DERs is that attacks can directly target the functionality of a component (e.g., switch battery off) or exploit the normal functionality to achieve an abnormal behavior (e.g., control battery to create load oscillation). Thus, this work suggests to extend the original CyPhERS methodology by considering and differentiating both target features that represent either the technical functionality or behavior of DER components. Whether a target feature is functional or behavioral depends on the model inputs (target feature lags and covariates). More precisely, even a model of the same feature can represent either the functionality or behavior depending on the input, as demonstrated in Fig. 7. In the depicted example, the attacker wants to create a sudden change of the PV feed. Instead of directly damaging the plant, the attacker uses the normal functionality (reduced feed if less sun) to launch the attack. The functionality model uses the local Ir measurement as input and predicts the expected feed reduction. Thus, no functional anomaly is flagged. The behavior model uses an external Ir measurement which results in a deviation between the predicted and actual feed. As a result, a behavioral anomaly is flagged. In case of only monitoring the functionality, the attack impact would not be detected. On the contrary, when solely monitoring the behavior, it could not be determined whether the anomaly stems from device disfunction or misuse.

Table 3 lists the extracted physical target features and associated model inputs for the PV-battery system demonstration case.[1] The average active load P_{fmean} is selected as target feature representing the functionality of PV1-4, BAT1-4, and M1-3. In these cases, model inputs exclusively stem from the component's variables and immediate inputs, or data of a redundant device. The behavior of the batteries is represented by $P^{\text{BAT}i}_{\text{bmean}}$, which is the average active load of the i-th battery modeled based on the time of the day and PV feed. Consequently, anomalies in $P^{\text{BAT}i}_{\text{bmean}}$ indicate abnormal battery behavior given the current time and PV feed. For the energy meters, the absolute sum of the active load readings $|P^{\text{M}i}_{\text{sum}}|$ modeled based on $P^{\text{M}i}_{\text{fmean}}$ is considered as behavioral feature. As the meters multicast on constant frequency, a sum/mean-mismatch can be a sign for abnormally many or few multicasts, potentially indicating misuse.

The battery operation in the given PV-battery system is influenced by weather and consumer behavior. The associated volatility and randomness potentially complicate the detection of anomalies. For such cases, this work proposes to extend the original CyPhERS methodology by extracting additional features that

Table 2: Raw network and process features considered in the demonstration case.

No.	Physical features	No.	Network features
1	Timestamp	1	Timestamp
2	Solar irradiation Ir	2-3	Internet protocol (IP) address
3-14	Act. power $P^{\text{PV1-4}}$, $P^{\text{BAT1-4}}$, $P^{\text{M1-4}}$	4-5	Media access control (MAC) address
15-18	Battery state of charge $SOC^{\text{BAT1-4}}$	6	Protocol
19-22	Battery voltage $V^{\text{BAT1-4}}$	7	TCP flags
23-26	Battery temperature $T^{\text{BAT1-4}}$	8	MB function code

[1]Note the restriction to features which can indicate load-altering events on a seconds to minutes scale due to the given resolution of the evaluated real data. Sophisticated attackers may also create sub-second power quality disturbances by modifying inverter output waveforms through firmware manipulation. Identifying the physical impact of such attacks would require extracting target features from high-frequency readings, such as the total harmonic distortion.

break down a component's complex behavior to simpler abstractions such as the on/off state. Two such abstracting features are considered for the demonstration case. $P^{\mathrm{BAT}i}_{\mathrm{osc}}$ describes how much power changes the i-th battery conducts in a 15s period. An unusually high value can be an indicator for abnormal load oscillation. Finally, $S^{\mathrm{BAT}i}$ describes the on/off state of a battery modeled under use of the current time of the day and PV feed. This target feature is supposed to indicate whether a battery is activated during times and PV feeds where it usually is deactivated and vice versa.

4.1.2. Network target features

Following the descriptions in [11], the purpose of traffic monitoring is twofold: 1) localization of compromised network devices, and 2) determination of attack vectors. To achieve the former, the traffic of each network device should be monitored individually. The latter necessitates extracting several informative features per device. Consequently, this work considers multiple network features for the PV and battery inverters, energy meters, DM, and DS (see Table 4). The features comprise counts of packets with specific protocols, TCP flags and MB function codes for periods of 15s. Within the local network of the considered PV-battery system, a variety of protocols are used, which enable certain functionalities, such as communicating process-relevant data (UDP packets from energy meters) or sending control commands (MB packets to battery inverters). Thus, unusual deviations in the packet count of certain protocols can point to specific attack types. Abnormal numbers of packets with SYN flags can be an indicator for adverse connection attempts and is thus taken into account. Finally, packets with function code 4 (read registers) and 16 (write registers) are counted. In particular, abnormal high numbers of packets with function code 16 can be a sign for adverse control commands. In all cases, network target features are modelled using lag values and the time of day as model input. Time of day is an important covariate in OT networks, as certain processes often are regularly conducted at specific times, e.g. every full hour.

Note that a broad spectrum of further useful features can be extracted. For example, the number of ports used within a certain period would provide information to identify port scans. Other features could be derived, among others, from IP/MAC addresses, packet sizes or checksums. However, for the sake of comprehensibility of the results in Section 5, only features which are considered most relevant are taken into account. Due to the same reason, only packets sent *to* a device are taken into consideration in most cases. Exceptions are M1-3 and the DS. As the meters M1-3 multicast process-relevant data, monitoring UDP packets send *from* them is of importance. Moreover, as the DS is available for users and thus directly connected to the external network, it constitutes a likely target for attackers. Thus, both packets send to and from the DS are monitored.

4.2. Signature extraction system

The signature extraction system generates anomaly flags for the set of target features, which eventually form the event signatures (see Fig. 2). The authors of CyPhERS argue for using individual models for each target feature [11]. Consequently, the signature extraction system comprises a set of individual anomaly detection and classification pipelines. Among the reasons is the independent selection of covariates, which becomes particularly relevant in context of the previously introduced differentiation between functionality- and behavior-describing target features.

The methodology of the anomaly detection and classification pipelines is explained in Section 4.2.1 with a particular focus on the transformation from deterministic to probabilistic models and detection rules as proposed in this work. After that, Section 4.2.2 addresses the time series models which are used within the pipelines. Finally, the procedure for automated implementation of the signature extraction system and its realization for the given demonstration case is detailed in Section 4.2.3.

4.2.1. Anomaly detection and classification pipelines

As explained in [11], a pipeline comprises a target feature model and consecutive anomaly detector (see Fig. 8). While the model predicts the normal behavior of the respective target feature, the detector compares the predictions with the ground truth observations to decide whether to flag an anomaly. In the original version of CyPhERS, both the predictions and the detector's decision function are deterministic [11]. Due to the weather- and consumer-induced randomness and variability of DER operation, modeling of some features is subject to pronounced uncertainties, rendering anomaly detection more challenging. Thus, this work

Table 3: Overview of physical target features extracted for the PV-battery system demonstration case.

Target feature	Model input	Type	Description
$P^{\mathrm{PV}i}_{\mathrm{fmean}}$	Ir (local)	Functional	For every 60s time step τ_{60} the mean value is determined as average over the $N_{\tau_{60}}$ data packets carrying $P^{\mathrm{PV}i}$ within τ_{60} according to $P^{\mathrm{PV}i}_{\mathrm{mean},\tau_{60}} = \sum_{p=1}^{N_{\tau_{60}}} P^{\mathrm{PV}i}_{p}/N_{\tau_{60}}$. Anomalies in $P^{\mathrm{PV}i}_{\mathrm{fmean}}$ can indicate disfunction of the i-th solar panel string or its inverter.
$P^{\mathrm{BAT}i}_{\mathrm{fmean}}$	$P^{\mathrm{M}i}_{\mathrm{mean}}$, $SOC^{\mathrm{BAT}i}_{\mathrm{mean}}$, $V^{\mathrm{BAT}i}_{\mathrm{mean}}$, $T^{\mathrm{BAT}i}_{\mathrm{mean}}$	Functional	For every τ_{60} the mean value is determined as average over the $N_{\tau_{60}}$ data packets carrying $P^{\mathrm{BAT}i}$ within τ_{60} according to $P^{\mathrm{BAT}i}_{\mathrm{mean},\tau_{60}} = \sum_{p=1}^{N_{\tau_{60}}} P^{\mathrm{BAT}i}_{p}/N_{\tau_{60}}$. Anomalies in $P^{\mathrm{BAT}i}_{\mathrm{fmean}}$ can indicate a disfunction of the i-th battery stack or its associated inverter.
$P^{\mathrm{M1\text{-}3}}_{\mathrm{fmean}}$	Redundant measurement of M4	Functional	For every 5s time step τ_{5} the mean value is determined as average over the $N_{\tau_{5}}$ data packets carrying $P^{\mathrm{M}i}$ within τ_{5} according to $P^{\mathrm{M}i}_{\mathrm{mean},\tau_{5}} = \sum_{p=1}^{N_{\tau_{5}}} P^{\mathrm{M}i}_{p}/N_{\tau_{5}}$. Anomalies can indicate a disfunction of the i-th meter.
$P^{\mathrm{BAT}i}_{\mathrm{osc}}$	$P^{\mathrm{BAT}i}_{\mathrm{osc}}$ lag values	Behavioral (Abstracting)	For every 15s time step τ_{15} the absolute sum of power changes is calculated according to $P^{\mathrm{BAT}i}_{\mathrm{osc},\tau_{15}} = \sum_{f=0s}^{15s} \lvert P^{\mathrm{BAT}i}_{\mathrm{mean},\tilde{\tau}+f+1s} - P^{\mathrm{BAT}i}_{\mathrm{mean},\tilde{\tau}+f} \rvert$, where $\tilde{\tau}$ is the start time of τ_{15}. Anomalies in $P^{\mathrm{BAT}i}_{\mathrm{osc}}$ can indicate oscillation of the i-th battery.
$S^{\mathrm{BAT}i}$	$P^{\mathrm{PV1\text{-}4}}_{\mathrm{mean}}$, time of day	Behavioral (Abstracting)	For every τ_{60} the on-off state ($S^{\mathrm{BAT}i}_{\tau_{60}} \in [0, 1]$) is determined. Anomalies may indicate that BATi is unexpectedly online/offline given the current time of day and PV feed.
$P^{\mathrm{BAT}i}_{\mathrm{bmean}}$	$P^{\mathrm{PV1\text{-}4}}_{\mathrm{mean}}$, time of day	Behavioral	For every τ_{60} the mean value is determined as average over the $N_{\tau_{60}}$ data packets carrying $P^{\mathrm{BAT}i}$ within τ_{60} according to $P^{\mathrm{BAT}i}_{\mathrm{mean},\tau_{60}} = \sum_{p=1}^{N_{\tau_{60}}} P^{\mathrm{BAT}i}_{p}/N_{\tau_{60}}$. Anomalies can indicate abnormal behavior of BATi given the current time of day and PV feed.
$\lvert P^{\mathrm{M1\text{-}3}}_{\mathrm{sum}} \rvert$	$P^{\mathrm{M}i}_{\mathrm{mean}}$	Behavioral	For every τ_{5} the absolute sum is determined as $\lvert P^{\mathrm{M}i}_{\mathrm{sum},\tau_{5}} \rvert = \sum_{p=1}^{N_{\tau_{5}}} P^{\mathrm{M}i}_{p}$. Anomalies can indicate abnormal behavior of the i-th meter given the current $P^{\mathrm{M}i}_{\mathrm{mean}}$ (sending abnormally many or few packets carrying $P^{\mathrm{M}i}$).

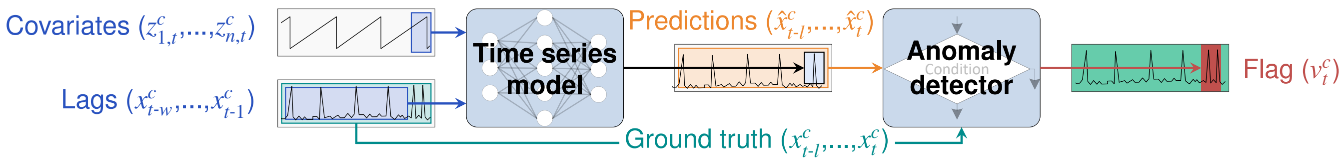


Figure 8: Schematic representation of the anomaly detection pipeline of a target feature c based on [11].

Table 4: Overview of network target features extracted for the PV-battery system demonstration case. Indices s and d refer to source and destination, respectively. Packets are counted for 15s periods.

Target feature	Description
$n^{\mathrm{PV}i}_{\mathrm{UDP_d}}, n^{\mathrm{BAT}i}_{\mathrm{UDP_d}}, n^{\mathrm{DM}}_{\mathrm{UDP_d}}, n^{\mathrm{DS}}_{\mathrm{UDP_d}}, n^{\mathrm{M}i}_{\mathrm{UDP_s}}, n^{\mathrm{DS}}_{\mathrm{UDP_s}}$	UDP packets send to/from the device
$n^{\mathrm{PV}i}_{\mathrm{TCP_d}}, n^{\mathrm{BAT}i}_{\mathrm{TCP_d}}, n^{\mathrm{DM}}_{\mathrm{TCP_d}}, n^{\mathrm{DS}}_{\mathrm{TCP_d}}, n^{\mathrm{DS}}_{\mathrm{TCP_s}}$	TCP packets send to/from the device
$n^{\mathrm{PV}i}_{\mathrm{MB_d}}, n^{\mathrm{BAT}i}_{\mathrm{MB_d}}, n^{\mathrm{DM}}_{\mathrm{MB_d}}, n^{\mathrm{DS}}_{\mathrm{MB_d}}, n^{\mathrm{DS}}_{\mathrm{MB_s}}$	MB packets send to/from the device
$n^{\mathrm{PV}i}_{\mathrm{ARP_d}}, n^{\mathrm{BAT}i}_{\mathrm{ARP_d}}, n^{\mathrm{DM}}_{\mathrm{ARP_d}}, n^{\mathrm{DS}}_{\mathrm{ARP_d}}, n^{\mathrm{DS}}_{\mathrm{ARP_s}}$	ARP packets send to/from the device
$n^{\mathrm{PV}i}_{\mathrm{TLS_d}}, n^{\mathrm{BAT}i}_{\mathrm{TLS_d}}, n^{\mathrm{DM}}_{\mathrm{TLS_d}}, n^{\mathrm{DS}}_{\mathrm{TLS_d}}, n^{\mathrm{DS}}_{\mathrm{TLS_s}}$	Transport layer security (TLS) packets send to/from the device
$n^{\mathrm{PV}i}_{\mathrm{SYN_d}}, n^{\mathrm{BAT}i}_{\mathrm{SYN_d}}, n^{\mathrm{DM}}_{\mathrm{SYN_d}}, n^{\mathrm{DS}}_{\mathrm{SYN_d}}, n^{\mathrm{DS}}_{\mathrm{SYN_s}}$	Packets with SYN flag send to/from the device
$n^{\mathrm{PV}i}_{16_\mathrm{d}}, n^{\mathrm{BAT}i}_{16_\mathrm{d}}$	Packets with write register code send to the device
$n^{\mathrm{PV}i}_{4_\mathrm{d}}, n^{\mathrm{BAT}i}_{4_\mathrm{d}}$	Packets with read register code send to the device

proposes to apply probabilistic time series prediction and decision functions when using CyPhERS for DER monitoring. In this way, the detection sensitivity is dynamically adjusted to the model's current confidence level, which potentially reduces the number of false positives during times of low confidence, and improves the detection of actual events at periods where model confidence is high. For that purpose, the lower quantile q_{L}, median q_{M}, and upper quantile q_{U} are predicted for each target feature. Given $X_c = \{x^c_1, x^c_2, ..., x^c_N \mid x^c_i \in \mathbb{R}\ \forall i\}$ and $Z_{c,1}, ..., Z_{c,n} = \{\{z^c_{1,1}, z^c_{1,2}, ..., z^c_{1,N}\}, ..., \{z^c_{n,1}, z^c_{n,2}, ..., z^c_{n,N}\} | z^c_{j,i} \in \mathbb{R} \forall (j,i)\}$ of a target feature c, the expected quantile $\hat{x}^{q,c}_t$ at time t is predicted using lag values $x^c_{t-w}, ..., x^c_{t-1}$ and covariates $z^c_{1,t}, ..., z^c_{n,t}$ according to

$$\hat{x}^{q,c}_t = \Phi\left([x^c_{t-w}, ..., x^c_{t-1}], [z^c_{1,t}, ..., z^c_{n,t}]\right), \forall q \in \{q_{\mathrm{L}}, q_{\mathrm{M}}, q_{\mathrm{U}}\}, \tag{1}$$

where n is the number of covariates, and w the length of the history window. Note that, depending on the target feature, $x^c_{t-w}, ..., x^c_{t-1}$ and $z^c_{1,t}, ..., z^c_{n,t}$ are only partially used (see Section 4.1). A model is trained to predict $\hat{x}^{q,c}$ by minimizing the quantile loss function [41]

$$L_q(\hat{x}^c_i, x^c_i) = \max\left[q(x^c_i - \hat{x}^c_i), (q-1)(x^c_i - \hat{x}^c_i)\right] \tag{2}$$

over a training set $X^c_{\mathrm{train}} = \{x^c_1, x^c_2, ..., x^c_{N_{\mathrm{train}}} \mid x^c_i \in \mathbb{R}\ \forall i\}$.

The quantile predictions of the target feature are forwarded to the anomaly detector. In the original version of CyPhERS, the detector decides whether to flag an anomaly based on the distance between ground truth observations and *deterministic* predictions [11]. The present work proposes to extend to the distance between the ground truth and the *probabilistic* prediction interval (PI) $[\hat{x}^{q_{\mathrm{L}},c}, \hat{x}^{q_{\mathrm{U}},c}]$. In this way, the dynamical adaption of the detection sensitivity to the current model's confidence is realized. When a model is certain about its predictions, even small deviations are accounted for. On the other hand, low model confidence (larger PI) will reduce the calculated distance. The distances are averaged over the last l observations according to

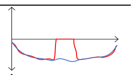
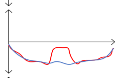
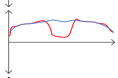
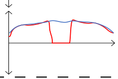
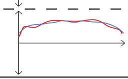
$$\varepsilon^c_t = \frac{\displaystyle\sum_{j=0}^{l-1} \begin{cases} x^c_{t-j} - \hat{x}^{q_{\mathrm{U}},c}_{t-j} & \text{if } x^c_{t-j} > \hat{x}^{q_{\mathrm{U}},c}_{t-j} \\ \hat{x}^{q_{\mathrm{L}},c}_{t-j} - x^c_{t-j} & \text{if } x^c_{t-j} < \hat{x}^{q_{\mathrm{L}},c}_{t-j} \end{cases}}{l}. \tag{3}$$

For the PV-battery system demonstration case, $l = 5$, $q_{\mathrm{L}} = 0.01$, and $q_{\mathrm{U}} = 0.99$ is selected $\forall c \in \mathcal{I}$ and $\mathcal{J}$. Based on ε^c_t and further characteristics of the current target feature observation, different anomaly types are distinguished, which is expressed by the decision function

$$v^c_t = \begin{cases} 2 & \text{if } \overbrace{(\varepsilon^c_t > \tau_c)}^{\text{Detection}},\ \overbrace{(x^c_t > \hat{x}^{q_{\mathrm{M}},c}_t)}^{\text{Direction}} \text{ and } \overbrace{(x^c_t = 0)}^{\text{Is zero}} \\ 1 & \text{if } (\varepsilon^c_t > \tau_c),\ (x^c_t > \hat{x}^{q_{\mathrm{M}},c}_t) \text{ and } (x^c_t \neq 0) \\ -1 & \text{if } (\varepsilon^c_t > \tau_c),\ (x^c_t < \hat{x}^{q_{\mathrm{M}},c}_t) \text{ and } (x^c_t \neq 0) \\ -2 & \text{if } (\varepsilon^c_t > \tau_c),\ (x^c_t < \hat{x}^{q_{\mathrm{M}},c}_t) \text{ and } (x^c_t = 0) \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

with τ_c being a target feature-specific threshold. The anomaly types are further explained in Table 5. Both the direction of an abnormally large deviation and the information about a target feature being zero provides valuable information for identification of event root causes and physical impact. For example, an abnormally *high* number of UDP packets send by an energy meter may indicate a FDIA, while a PV feed of *zero* during daytime may points towards a switched off inverter.

Table 5: Definition of considered anomaly types.

Flag v	Anomaly type	Description	Schematic
2	Positive zero	Target feature abnormally high and zero.	
1	Positive non-zero	Target feature abnormally high and non-zero.	
−1	Negative non-zero	Target feature abnormally low and non-zero.	
−2	Negative zero	Target feature abnormally low and zero.	
0	No anomaly	No abnormal behavior.	

4.2.2. Time series models

The authors of CyPhERS suggest the use of specific models for different target feature classes, which includes the use of long-short term memory (LSTM) networks for network traffic features. However, given the computational limitations of small DERs, the use of resource intensive models is impractical. Thus, this work proposes to apply gradient-boosted decision trees (GBDT) [42] for predicting the quantiles $\forall c \in \mathcal{I}$ and $\mathcal{J}$. GBDT is a frequently applied ML technique, popular for high accuracy and efficiency, which renders them a good fit for the given problem [43]. GBDT consists of a set of simple decision trees, which are connected in series. Thus, each of them minimizes the prediction error of the preceding tree. For a detailed explanation, the reader is referred to [42].

4.2.3. Automated model and detector tuning procedure

In [11], the authors of CyPhERS suggest an automated implementation procedure for the signature extraction system, which comprises independent tuning of the individual detection and classification pipelines (see Fig. 9). First step is the training and hyperparameter selection of the GBDT models of all target features. For the PV-battery system demonstration case, selection of hyperparameters is conducted on 75 % of the two-week normal operation data. The tuned hyperparameters and associated search spaces are summarized in Table 6.

Table 6: Hyperparameters and search spaces of the GBDT models.

No.	Hyperparameter	Search space
1	w^{a}	[0,...,60]
2	Max. depth of a tree	[3,...,21]
3	Number of decision trees	[100,...,1000]
4	Learning rate	[0.001,...,0.3]

[a]Only for target features considering lag values.

Next, the anomaly detectors of all pipelines are tuned. For that purpose, the fitted models are used to predict the remaining 25 % of normal operation data. Based on the predictions, the distances $E^c_{\text{test}} = \{\varepsilon^c_1, \varepsilon^c_2, ..., \varepsilon^c_{N_{\text{test}}} \mid \varepsilon^c_i \in \mathbb{R}\ \forall i\}$ are calculated according to (3), $\forall c \in \mathcal{I}$ and $\mathcal{J}$. From E^c_{test} and a threshold factor f, the feature-specific thresholds are determined according to

$$\tau_c = f \cdot \max\left(E^c_{\text{test}}\right), \tag{5}$$

where the threshold factor is selected as $f = 1.1$ in this work. Therefore, an anomaly within a target feature c at time t is flagged if ε^c_t exceeds the largest distance during normal operation by at least 10 %. Before the anomaly detection pipelines are applied for online flagging of new observations, the models are retrained on the entire set of historical normal operation data (see Fig. 9).

Finally, the resulting anomaly flag series provided by the individual pipelines are grouped for each system zone to obtain human-readable event signatures as output of CyPhERS' Stage 1. For the PV-battery system demonstration case, the defined system zones comprise PV1-4, BAT1-4, M1-3, DM, and DS.

4.3. Signature evaluation (Stage 2)

This section details the methodology of CyPhERS' signature evaluation with a focus on the realization of an automated signature evaluation system, and describes its implementation for the PV-battery system demonstration case. First, the definition of a signature database is addressed in Section 4.3.1, followed by an explanation of the proposed automated signature evaluation system in Section 4.3.2.

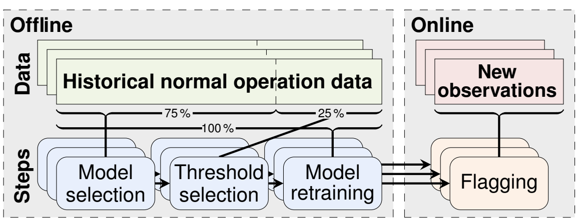


Figure 9: Procedure for tuning the anomaly detection pipelines based on [11].

4.3.1. Signature database

Signature evaluation in CyPhERS is based on manually or automatically matching the organized anomaly flags of the set of target features provided by Stage 1 with a database of known event signatures [11]. The signatures of the attack types included in the demonstration case are depicted in Fig. 10 on the example of selected victim devices and physical impacts. The associated signature descriptions are provided in Table 7. For the sake of conciseness, target features of the same kind are jointly represented in one row in Fig. 10. More precisely, for individual system zones, the various protocol count flags are combined in $v_{n_{\text{proto}}}$, and flags in counts of different MB function codes ($v_{n_{16_d}}$ and $v_{n_{4_d}}$) in $v_{n_{\text{fcode}}}$. Signatures can also be defined for unknown event types, which then carry reduced information such as indication of affected system zones. For the sake of clarity, however, these are not included in Fig. 10.

4.3.2. Automated signature evaluation

The demonstration of the original version of CyPhERS in [11] includes manual recognition and evaluation of the event signatures provided by Stage 1. In case of larger plants, such as medium voltage level-connected wind and solar parks, visual recognition of event signatures in control rooms may be feasible. However, for smaller resources such as residential PV-battery systems, manual evaluation is impractical. For the sake of proving feasibility of automated signature evaluation, this work formulates and applies a rule-based signature evaluation system which jointly evaluates the most recent flags of all target features within a moving

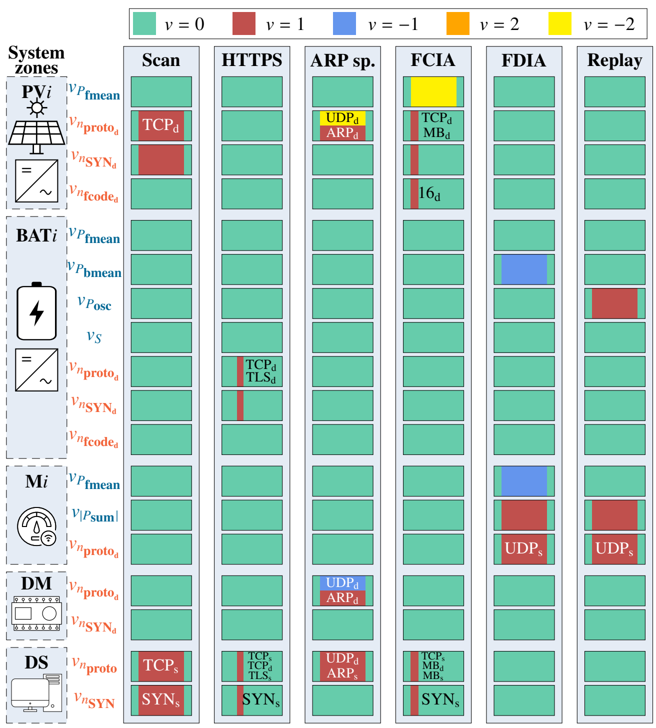


Figure 10: Event signatures of the attack types included in the demonstration case. The signatures are depicted for selected victim devices and physical impacts.

Table 7: Description of the event signatures of the attack types included in the demonstration case (see Fig. 10).

Attack	Event signature description
Scan	A device (e.g., PV inverter) receives an abnormally large number of TCP packets ($v_{n_{\mathrm{TCP_d}}} = 1$) with connection request ($v_{n_{\mathrm{SYN_d}}} = 1$) over a longer period. Simultaneously, another device (e.g., DS) sends unusually many TCP packets ($v_{n_{\mathrm{TCP_s}}} = 1$) with connection request ($v_{n_{\mathrm{SYN_s}}} = 1$). Together, this points towards scanning of a victim device (here, the PV inverter), where the attacker is located on a local device (here, the DS). The lack of anomaly flags in physical target features indicates a pure cyber attack without physical impact.
HTTPS request	A device (e.g., battery inverter) receives abnormally many TLS packets ($v_{n_{\mathrm{TLS_d}}} = 1$) over short period, pointing towards a web service call (HTTPS request). Simultaneously, another device (e.g., DS) sends more TLS packets than usual ($v_{n_{\mathrm{TLS_s}}} = 1$), which suggests the attacker being located on this device. Parallel increase of TCP packets ($v_{n_{\mathrm{TCP_d}}} = 1$) and packets with SYN flags ($v_{n_{\mathrm{SYN_d}}} = 1$) due to connection establishment between attacker and victim device. The absence of anomaly flags in physical target features indicates a cyber attack without any physical impact.
ARP spoof	Two devices (e.g., PV inverter and DM) receive abnormally many ARP packets ($v_{n_{\mathrm{ARP_d}}} = 1$), while another (e.g., DS) sends more than expected ($v_{n_{\mathrm{ARP_s}}} = 1$). This points towards ARP spoofing where the attacker is located on a local device (here, the DS). The two victim devices receive less (or no) UDP packets ($v_{n_{\mathrm{UDP_d}}} = -1$ or -2), while the device occupied by the attacker receives more ($v_{n_{\mathrm{UDP_d}}} = 1$), which suggests that the communication between the victims is successfully redirected via the occupied device[a]. Lack of flags in physical features imply eavesdropping instead of manipulation of process-relevant data.
FCIA	A device (e.g., PV inverter) receives an abnormally large number of MB packets ($v_{n_{\mathrm{MB_d}}} = 1$) with write register function code 16 ($v_{n_{16_\mathrm{d}}} = 1$). In parallel, another device (e.g., DS) sends more MB packets than usual ($v_{n_{\mathrm{MB_s}}} = 1$). Together, this indicates an attacker sending false control commands to a victim device (here, the PV inverter) from the occupied local device (here, the DS). Parallel increase of TCP packets and packets with SYN flags because of connection establishment between occupied and victim device. Abnormally low and zero PV feed ($v_{P_{\mathrm{fmean}}} = -2$) indicate that the attacker switched off the PV inverter[b].
FDIA	An energy meter Mi sends unusually many UDP packets[c] ($v_{n_{\mathrm{UDP_s}}} = 1$) while the absolute sum of its active power readings is too high ($v_{\lvert P_{\mathrm{sum}} \rvert} = 1$). Together this points towards unusual frequent broadcasting of active power readings. The parallel abnormally low mean ($v_{P_{\mathrm{fmean}}} = -1$) indicates false $P^{\mathrm{M}i}$ injection imitating grid exports. For the battery which uses Mi readings, an unusually low mean active power given the current time and PV feed ($v_{P_{\mathrm{bmean}}} = -1$) suggests reaction with charging[b]. Absence of anomalies in $P^{\mathrm{BAT}i}_{\mathrm{fmean}}$ underlines that the battery accepts the false data and reacts to them in an expected way.
Replay attack	An energy meter Mi sends abnormally high numbers of UDP packets[c] ($v_{n_{\mathrm{UDP_s}}} = 1$), and the absolute sum of its active power measurements is higher than expected ($v_{\lvert P_{\mathrm{sum}} \rvert} = 1$). Together this indicates unusually frequent broadcasting of active power readings. As the mean is normal ($v_{P_{\mathrm{fmean}}} = 0$), no false data is injected, and instead, a replay of valid $P^{\mathrm{M}i}$ readings is likely. Abnormally high power changes ($v_{P_{\mathrm{osc}}} = 1$) of one or more batteries indicates load oscillation due to multiplication of the control error through replaying $P^{\mathrm{M}i}$ values.

[a]Parallel network anomalies for other devices which communicate with the victims possible as victim functionality can be affected by the attack.
[b]Physical impact depends on the malicious control command/injected false data, and the victim device.
[c]Parallel network anomalies for other devices possible due to UDP traffic overloading of those.

time window T_{eval} of five minutes. A simplified representation is provided in Algorithm 1. In case that no anomaly is detected (i.e., $v_c = 0$, $\forall c \in \mathcal{I}$ and $\mathcal{J}$) within T_{eval}, the system predicts normal operation. Given that the flags within T_{eval} match with one of the pre-defined event signatures, the associated event description forms the prediction, for example, *FCIA against PV1 from DS with physical impact "PV1 switched off"*. In case that anomaly flags provided by Stage 1 do not match any of the defined rules, the evaluation system predicts *Unknown abnormal behavior*. In contrast to the simplified representation in Algorithm 1, additional rules for different variations of specific attack types are defined. For example, *Replay attack against meter M1 with physical impact "oscillation" on BAT1* can be *with* or *without parallel traffic overloading of other devices*. Further rules predict affected system zones in case of unknown event types, for example, *Unknown network anomaly affecting PV2 and DS*. Finally, some rules additionally take previous predictions into account. For example, in case of predicting a battery being switched off by a FCIA, information about the injected false command is stored over a time period larger than T_{eval} in order to avoid switching to prediction of a pure physical failure after five minutes.

5. Demonstration of CyPhERS for DER monitoring

This section first demonstrates results from applying the adapted version of CyPhERS to the experimentally derived dataset of a real PV-battery system, as introduced in Section 3. The included attack types are successively evaluated in Sections 5.1-5.6. Thereafter, a quantitative performance comparison to the original version of CyPhERS [11] is conducted in Section 5.7 to assess the impact of the proposed methodological adaptations. The binary (*normal* vs. *abnormal operation*) detection performance is compared based on the true and false positive rate according to

$$TPR = \frac{N_{\mathrm{TP}}}{N_{\mathrm{P}}} \text{ and } FPR = \frac{N_{\mathrm{FP}}}{N_{\mathrm{N}}}, \tag{6}$$

where N_{TP}, N_{FP}, N_{P}, and N_{N} are the number of true positive, false positive, actual positive, and actual negative observations, respectively. The binary anomaly flags are created following

$$v_t^{\mathrm{binary}} = \begin{cases} 1 \ (\textit{abnormal operation}) & \text{if } v_t^c \neq 0, \ \exists c \in \mathcal{J} \cup \mathcal{I} \\ 0 \ (\textit{normal operation}) & \text{otherwise.} \end{cases} \tag{7}$$

Observations within a 5-minute window after each attack event are excluded from the *FPR* calculation to avoid considering positive anomaly flags ($v_t^{\mathrm{binary}} = 1$) during recovery of the attacked system as false positives, which would bias the performance assessment.

The identification performance is evaluated for each attack type a based on the identification rate according to

$$IR_a = \frac{N_a^{\mathrm{identified}}}{N_a}, \tag{8}$$

with N_a being the total number of event instances of attack type a and $N_a^{\mathrm{identified}}$ the corresponding subset of correctly identified instances.

5.1. Scanning attacks

The event signatures provided by CyPhERS' Stage 1 during the two scanning attacks are depicted in Fig. 11 together with the predictions of the rule-based signature evaluation system (Stage 2). Note that system zones without flagged anomalies are not depicted in the following. In both cases the provided signatures correspond to the signature of scanning attacks (see Fig. 10). Thus, visual recognition of the signature allows identifying the attack type (scan), victim (PV3 or PV4, respectively), and attacker location (DS) manually. The same predictions are provided by the rule-based system without human interaction. During the first scan, the rule for predicting a scan is not immediately fulfilled, since flagging $v_{n^{\mathrm{DS}}_{\mathrm{TCP_s}}} = 1$ is delayed. Therefore, the rule-based system initially predicts an unknown network traffic anomaly for PV3 and DS, based on the occurrence of flags in the associated network target features.

Fig. 12 exemplifies the advantage of modeling target features with time series models. Since $n^{\mathrm{DS}}_{\mathrm{TCP_s}}$ exhibits normal peaks at full hours, the increase during scanning of PV4 only constitutes a local anomaly which cannot be detected by static thresholds not taking temporal information into account. In contrast, the applied GBDT model allows to detect the scanning-induced local anomaly by learning that peaks should only occur at full hours.

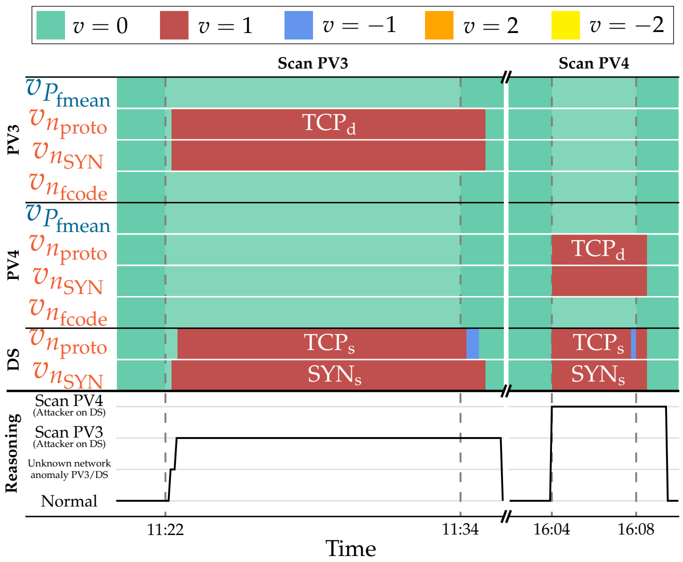


Figure 11: Event signatures provided by CyPhERS' Stage 1 and predictions of the rule-based signature evaluation system (Stage 2) during the scan attacks.

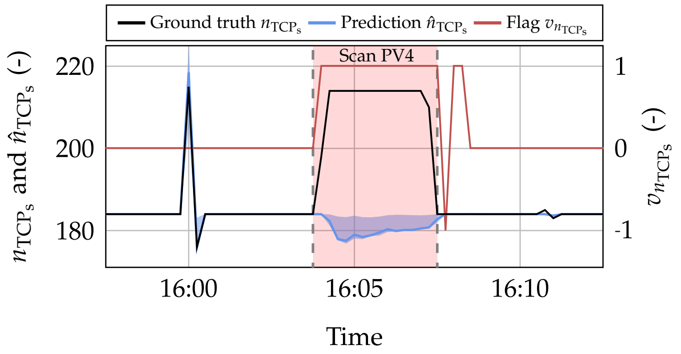


Figure 12: Ground truth, prediction (98 % PI and median) and anomaly flag for $n^{\mathrm{DS}}_{\mathrm{TCP_s}}$ during scan of PV4.

5.2. *HTTPS request attacks*

The provided event signatures and rule-based predictions for the two HTTPS requests are depicted in Fig. 13. Due to a match with the signature of HTTPS requests (see Fig. 10), the attack type can be identified together with the victims (BAT1 and DM, respectively), and attacker location (DS), both through visual signature recognition and the rule-based system.

5.3. *ARP spoofing attacks*

Fig. 14 depicts the provided event signatures and predictions of the rule-based system for both ARP spoofing attacks. The signatures match the one for ARP spoofing (see Fig. 10). Since the associated rules are fulfilled, Stage 2 predicts ARP spoofing attacks from an attacker located on the DS against PV3|DM, and PV4|DM, respectively. As the attacks distract the DM, its UDP communication pattern to non-victim devices is also affected, resulting in parallel network anomaly flags for BAT1 during the first ARP spoof and PV3 during the second. As this behavior is considered as a sub-case of the ARP spoofing signature, and integrated as such in the rule-based system, predictions switch between ARP spoofing with and without parallel traffic distraction of other devices (see Fig. 14).

Algorithm 1 Simplified representation of the proposed rule-based signature evaluation system.

$T_{\mathrm{eval}} \leftarrow$ Last 5 minutes
if all flags in T_{eval} are zero **then**
  prediction $\leftarrow$ Normal operation
else if $v^X_{n_{\mathrm{TCP_d}}} = 1$, $v^X_{n_{\mathrm{SYN_d}}} = 1$, $v^Y_{n_{\mathrm{TCP_s}}} = 1$, and $v^Y_{n_{\mathrm{SYN_s}}} = 1$ within T_{eval} **then**
  prediction $\leftarrow$ Scan of device X from device Y
else if $v^X_{n_{\mathrm{TLS_d}}} = 1$, $v^X_{n_{\mathrm{TCP_d}}} = 1$, $v^X_{n_{\mathrm{SYN_d}}} = 1$, $v^Y_{n_{\mathrm{TLS_s}}} = 1$, $v^Y_{n_{\mathrm{TCP_d}}} = 1$, $v^Y_{n_{\mathrm{TCP_s}}} = 1$, and $v^Y_{n_{\mathrm{SYN_s}}} = 1$ within T_{eval} **then**
  prediction $\leftarrow$ HTTPS request of device X from device Y
else if $v^X_{n_{\mathrm{ARP_d}}} = 1$, $v^Y_{n_{\mathrm{ARP_d}}} = 1$, $v^Z_{n_{\mathrm{ARP_s}}} = 1$, $v^X_{n_{\mathrm{UDP_d}}} = -1$ or -2, $v^Y_{n_{\mathrm{UDP_d}}} = -1$ or -2, and $v^Z_{n_{\mathrm{UDP_d}}} = 1$ within T_{eval} **then**
  prediction $\leftarrow$ ARP spoof against devices X,Y from device Z
else if $v^X_{n_{\mathrm{TCP_d}}} = 1$, $v^X_{n_{\mathrm{MB_d}}} = 1$, $v^X_{n_{\mathrm{SYN_d}}} = 1$, $v^X_{n_{16_d}} = 1$, $v^Y_{n_{\mathrm{TCP_s}}} = 1$, $v^Y_{n_{\mathrm{MB_d}}} = 1$, $v^Y_{n_{\mathrm{MB_s}}} = 1$, $v^Y_{n_{\mathrm{SYN_s}}} = 1$, and $v^X_{P_{\mathrm{fmean}}} = -2$ within T_{eval} **then**
  prediction $\leftarrow$ FCIA against device X from device Y with physical impact A (here, switch X off)
else if $v^M_{n_{\mathrm{UDP_s}}} = 1$, $v^M_{|P_{\mathrm{sum}}|} = 1$, $v^M_{P_{\mathrm{fmean}}} = -1$, and $v^X_{P_{\mathrm{bmean}}} = -1$ within T_{eval} **then**
  prediction $\leftarrow$ FDIA against meter M with physical impact A on device X (here, battery charging)
else if $v^M_{n_{\mathrm{UDP_s}}} = 1$, $v^M_{|P_{\mathrm{sum}}|} = 1$, $v^M_{P_{\mathrm{fmean}}} = 0$, and $v^X_{P_{\mathrm{osc}}} = 1$ within T_{eval} **then**
  prediction $\leftarrow$ Replay attack against meter M with physical impact A on device X (here, battery oscillation)
else
  prediction $\leftarrow$ Unknown abnormal behavior

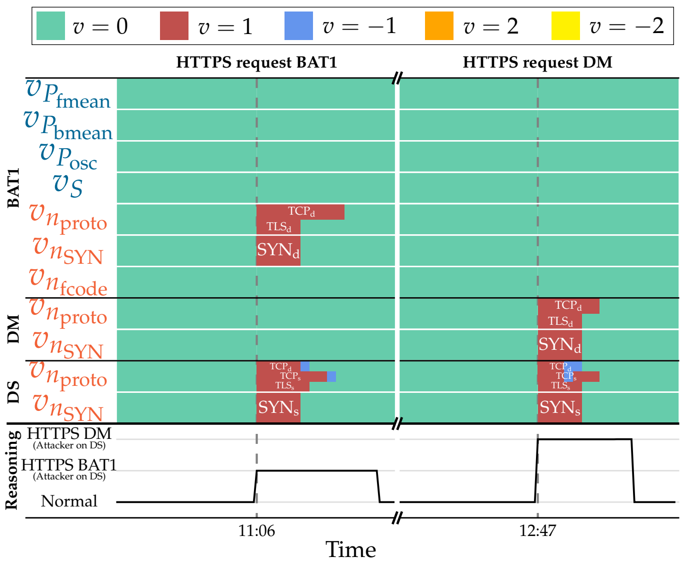


Figure 13: Event signatures provided by CyPhERS' Stage 1 and predictions of the rule-based signature evaluation (Stage 2) during the HTTPS requests.

Another event is detected shortly before the second ARP spoof. As the provided anomaly flags do not match with the signature of a known attack, reduced event information (occurrence, affected network device, no physical impact) is provided by Stage 2. This example demonstrates that CyPhERS can automatically provide information such as event occurrence and affected devices also for unknown event types.

To illustrate the prediction and flagging process of the underlying anomaly detection and classification pipelines, some examples are provided in Fig. 15. Fig. 15 (a) represents $n^{\mathrm{PV4}}_{\mathrm{ARP_d}}$ during the second ARP spoofing attack. It can be noticed that the spoofs result in pronounced global anomalies which are immediately detected. As ARP packets during normal operation occur non-deterministically, the small peaks cannot be learned by the GBDT model. Instead, it puts the PI on a constant level to capture those peaks, which illustrates that the GBDT model approximates a static but accurate threshold in cases without learnable pattern. Fig. 15 (b) shows $n^{\mathrm{DM}}_{\mathrm{UDP_d}}$ during the first ARP spoof. As the DM maintains UDP communication with non-victim devices, the oscillation pattern persists, however, on a lower level. The level decrease is detected by the underlying pipeline. Fig. 15 (c) depicts an excerpt of the UDP traffic distraction of BAT1 during the first ARP spoof. It can be seen that the distraction only expresses as a local and short pattern interruption without specific traffic increase or decrease.

5.4. *False command injection attacks*

Fig. 16 depicts the event signatures of Stage 1 and predictions of Stage 2 during the four FCIAs. In all cases, the provided signatures match the FCIA signature (see Fig. 10), allowing to identify the attack type, victims, attacker location, and physical impact, as the rule-based predictions indicate. Fig. 17 illustrates the detection of false commands on the example of $n^{\mathrm{DS}}_{\mathrm{MB_s}}$ during the FCIA against PV1. The underlying GBDT model successfully learned the normal peaks at full hours. Moreover, it understands that small positive peaks are usually followed by negative ones. As the injection of false commands is not followed by a negative peak, the larger distance between prediction and ground truth results in an anomaly flag.

The yellow or orange flags ($v = -2$ or 2) in the physical target features indicate that the attacker switched off the respective victim device. The detection of the physical impact is exemplified on the FCIAs against PV1 and BAT3[2] in Fig. 18. It can be noticed that modeling of physical target features is subject to larger uncertainties compared to network traffic modeling. As a result, smaller physical impacts may be missed by some features as, for example, the case for $v_{P^{\mathrm{BAT4}}_{\mathrm{fmean}}}$ and $v_{P^{\mathrm{BAT4}}_{\mathrm{bmean}}}$ during the FCIA against BAT4. Note that, on the abstraction level of the battery state S, the switch-off is detected, which highlights the importance of such abstracting target features for applying CyPhERS for DER monitoring, as proposed in this work.

[2] Note that the predicted sudden change from charging to discharging in Fig. 18 (b) results from the compressor load peak that the battery would compensate if not switched off.

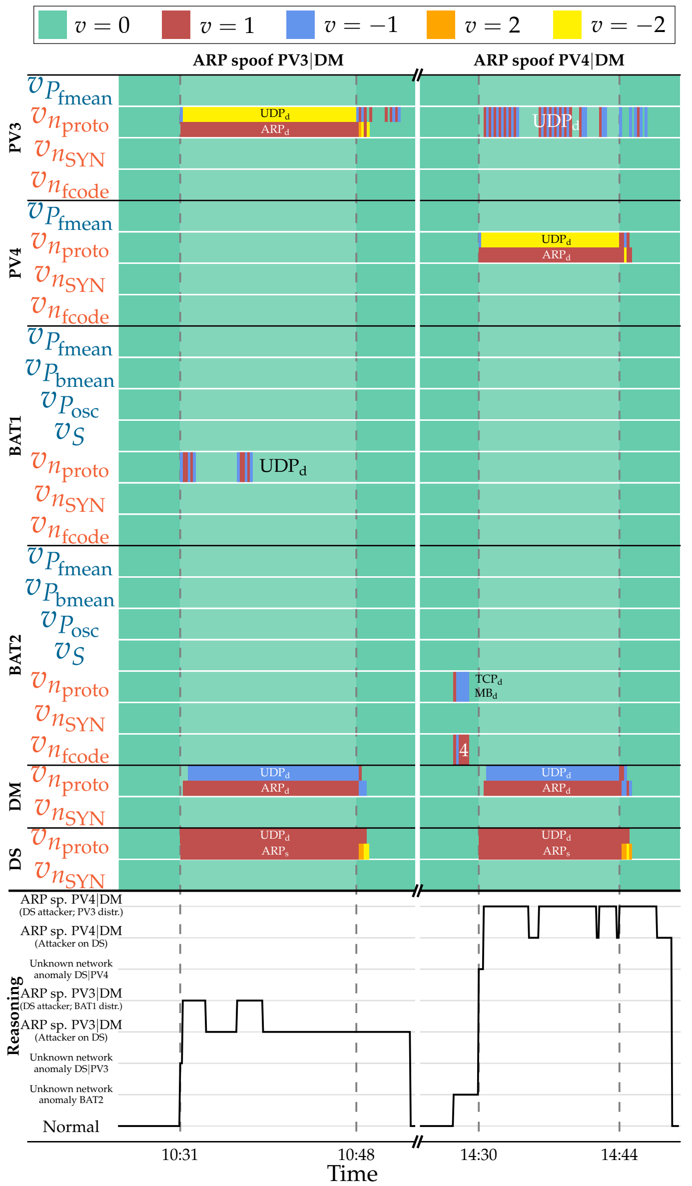


Figure 14: Event signatures provided by CyPhERS' Stage 1 and predictions of the rule-based signature evaluation system (Stage 2) during the ARP spoofs.

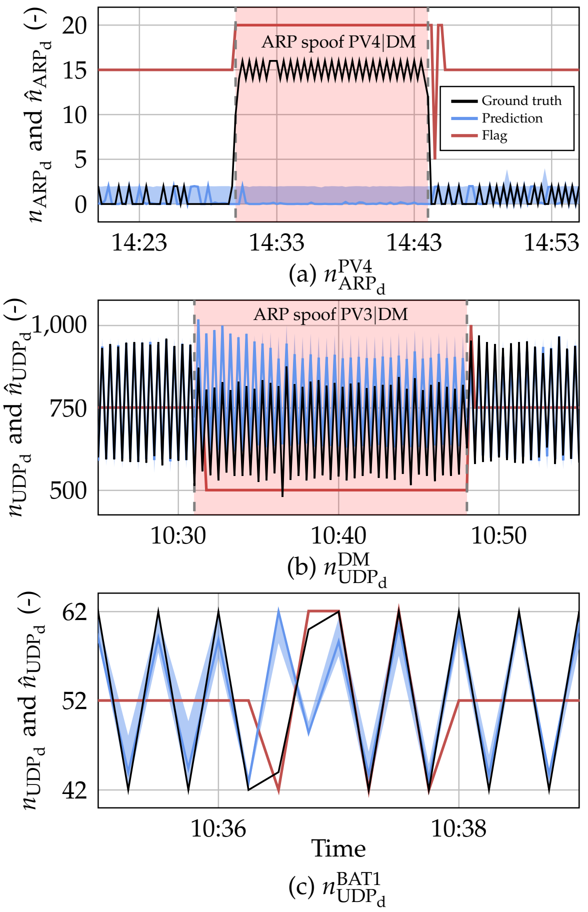


Figure 15: Ground truth, prediction (98 % PI and median), and anomaly flag for (a) $n_{\mathrm{ARP_d}}^{\mathrm{PV4}}$ during second ARP spoof, (b) $n_{\mathrm{UDP_d}}^{\mathrm{DM}}$ during first ARP spoof, and (c) $n_{\mathrm{UDP_d}}^{\mathrm{BAT1}}$ during first ARP spoof.

5.5. *False data injection attacks*

Fig. 19 depicts the provided event signatures of Stage 1 and predictions of the rule-based system (Stage 2) during the FDIAs. As the signatures during all three attacks correspond to the ones of FDIAs (see Fig. 10), attack type, victim device, and physical impact can be derived through visual signature recognition or automated rule-based predictions. Since the batteries accept the injected false data and react to them in an expected way, no disfunctionality is flagged ($v_{P_{\mathrm{fmean}}^{\mathrm{BAT}i}} = 0$). At the same time, anomaly flags in $v_{P_{\mathrm{bmean}}^{\mathrm{BAT}i}}$ indicate untypical battery behavior given the current time of the day and PV feed[3]. While blue flags ($v_{P_{\mathrm{bmean}}} = -1$) indicate abnormal charging, red flags ($v_{P_{\mathrm{bmean}}} = 1$) point towards unusual discharging. This example underlines the importance of considering both functional and behavioral target features for identification of the physical attack impact in case of DER monitoring, as suggested in this work.

5.6. *Replay attacks*

The event signatures of Stage 1 and predictions of Stage 2 during the two replay attacks are depicted in Fig. 20. During both attacks, anomalies are flagged in almost all system zones as the network devices are distracted by processing the large number

[3]Note that $v_S^{\mathrm{BAT3}} = 1$ during the first FDIA indicates that BAT3 was first activated by the attack.

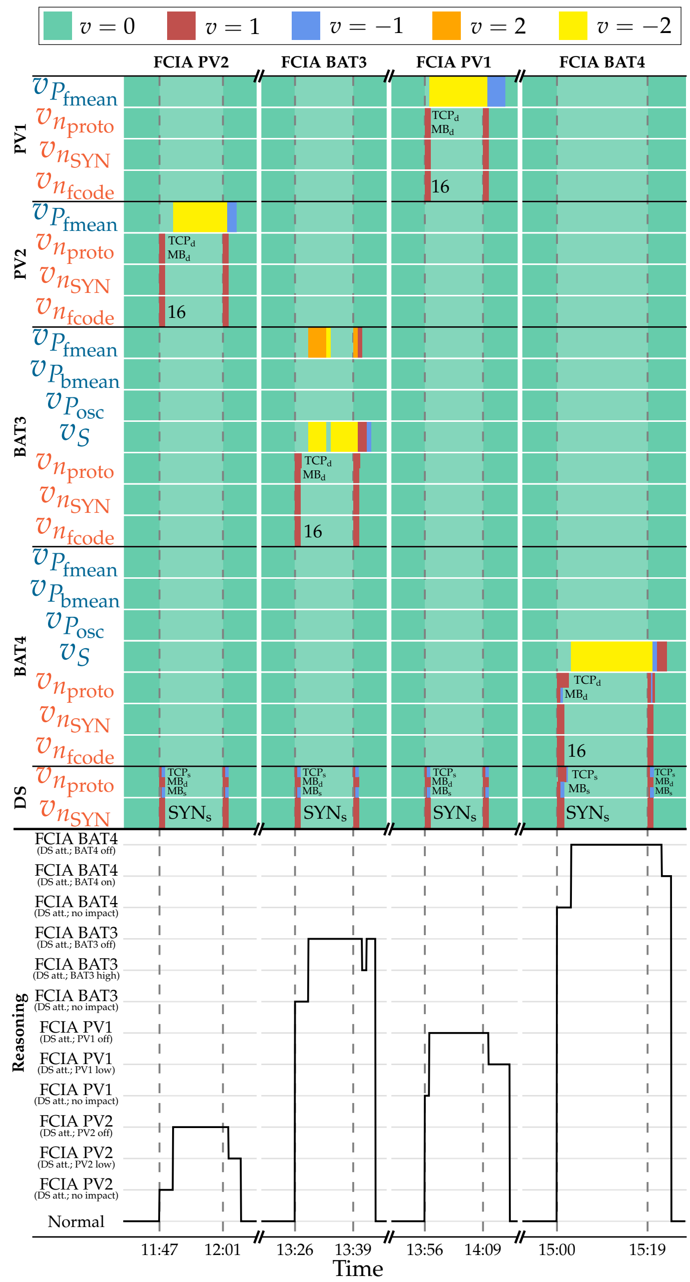


Figure 16: Event signatures provided by CyPhERS' Stage 1 and predictions of the rule-based signature evaluation system (Stage 2) during the FCIAs.

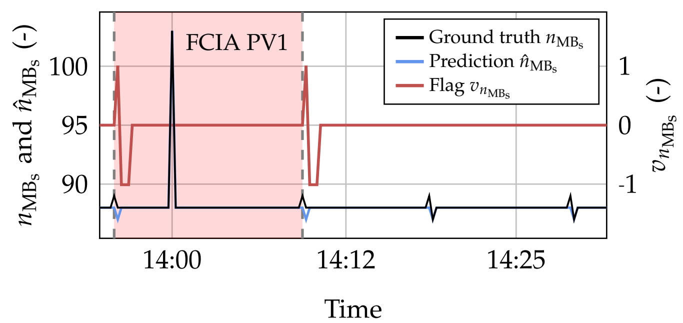


Figure 17: Ground truth, prediction (98 % PI and median) and anomaly flag for $n_{\mathrm{MB_s}}^{\mathrm{DS}}$ during FCIA against PV1.

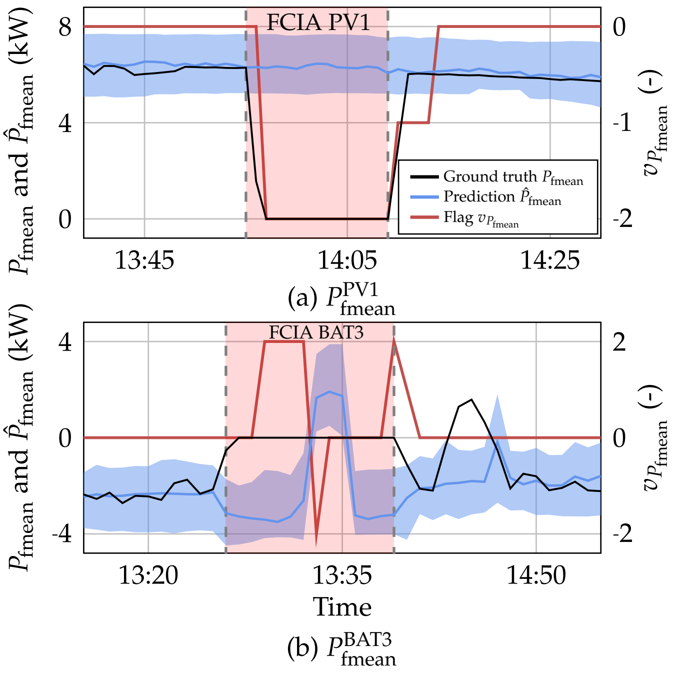


Figure 18: Ground truth, prediction (98 % PI and median) and anomaly flag for (a) $P^{\mathrm{PV1}}_{\mathrm{fmean}}$ during FCIA against PV1, and (b) $P^{\mathrm{BAT3}}_{\mathrm{fmean}}$ during FCIA against BAT3.

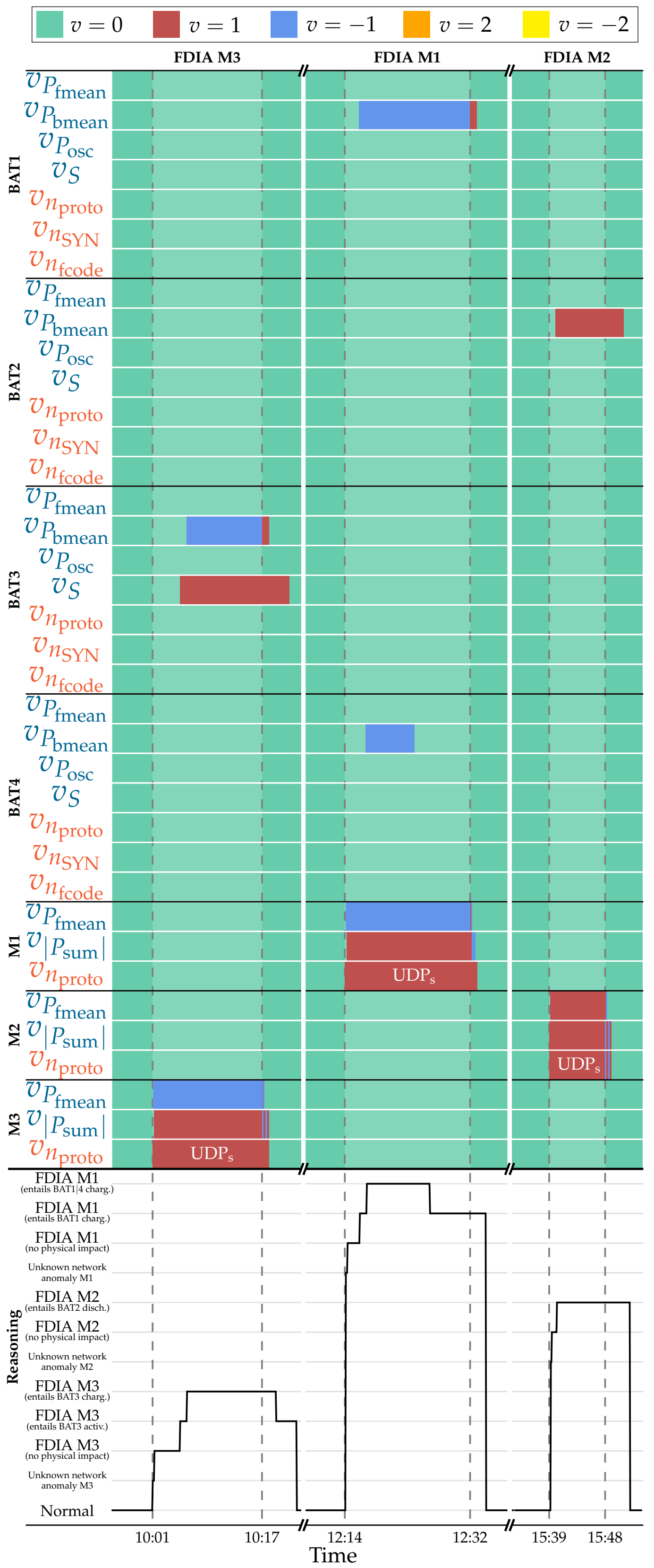


Figure 19: Event signatures provided by CyPhERS' Stage 1 and predictions of the rule-based signature evaluation system (Stage 2) during the FDIAs.

of replayed energy meter multicasts. In this case, visual recognition of specific attack patterns is challenging. In contrast, the rule-based system quickly identifies the signature as the associated rules are still fulfilled. Since parallel traffic flooding is integrated into the rule-based system as a sub-case of the replay attack signature, Stage 2 predicts the attack type, victim device, and physical impact along with network flooding. The correct identification of affected batteries and the physical impact on them is the result of incorporating abstracting target features, as proposed in this work (load oscillation of BATi indicated by $v_{P^{\mathrm{BAT}i}_{\mathrm{osc}}} = 1$).

The batteries which are controlled based on the replayed power readings oscillate between full charging and discharging power. Since they reach their maximum power limits, the other batteries take over, as explained in Section 3.1, and thus, begin to oscillate as well. However, since BAT1 and BAT4 are fully discharged during the second attack, no oscillation is indicated for them.

Towards the end of the second replay attack, the inverter in PV3 crashes since it cannot process the large number of packets as pointed out by the yellow flags in the network target features. Shortly after, $v_{P^{\mathrm{PV3}}_{\mathrm{fmean}}} = -2$ indicates that also the feed of the associated solar panel string is interrupted.

5.7. Performance impact of the methodological adaptations

Table 8 provides a performance comparison of applying the original and adapted version of CyPhERS to the dataset of the considered PV-battery system demonstration case. The adapted version achieves a significantly lower false positive rate. This improvement is mainly caused by the switch to *probabilistic models and detection rules*, which enable automatic reduction of the detection sensitivity at times of low confidence of the prediction model due to randomness and volatility in DER operation, thus, reducing false positives. From the identification rates in Table 8, it can be further seen that the identification of cyber attacks without physical impact (Scan, HTTPS, and ARP spoof) can be achieved based on the event signatures of both CyPhERS versions with comparable performance. However, the original CyPhERS fails to provide informative signatures for the identification of the cyber-

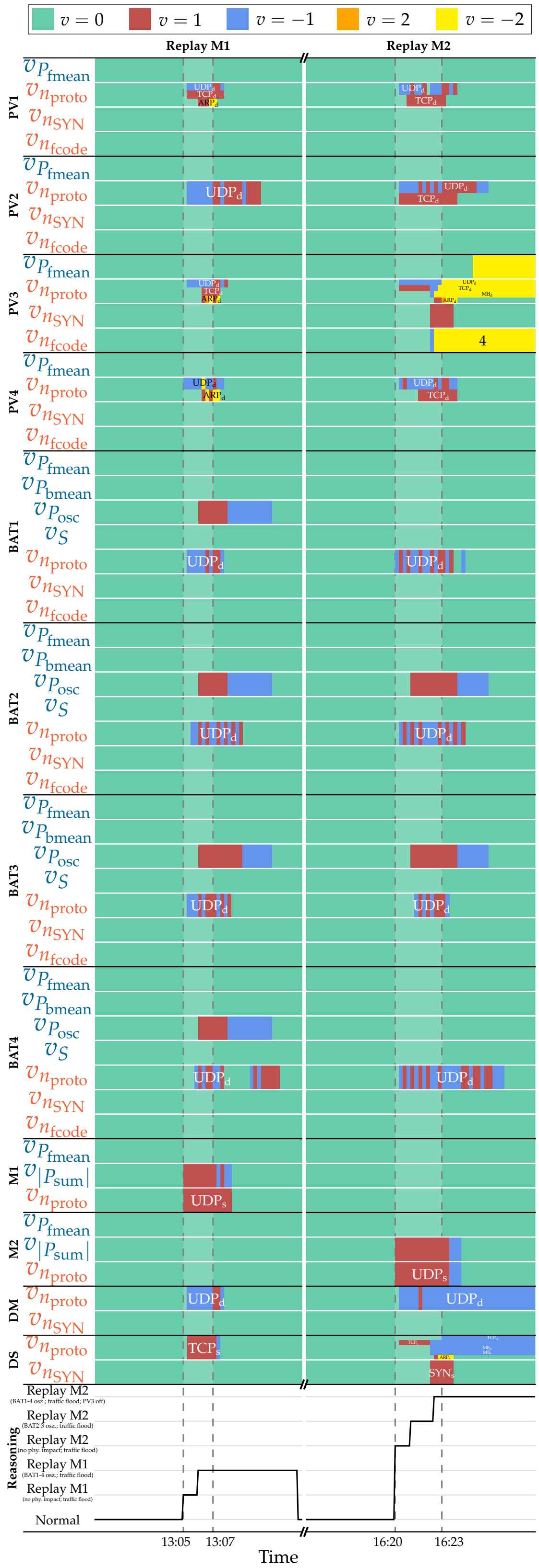


Figure 20: Event signatures provided by CyPhERS' Stage 1 and predictions of the rule-based signature evaluation system (Stage 2) during the replay attacks.

physical attacks (FCIA, FDIA, and replay attack) in most cases. The FCIAs against BAT3 and BAT4 cannot be identified as the abnormal battery operation is not detected and described in the provided signatures, which is due to two reasons: 1) The lower sensitivity of the deterministic detection approach at times of high model confidence, and 2) the lack of *abstracting physical target features* that break down modeling complexity of the battery operation, such as the on/off state of a battery. Finally, the FDIAs and replay attacks cannot be identified by evaluating the event signatures of the original CyPhERS version due to the lack of *behavioral target features* which indicate misuse of the normal battery functionality, for example, oscillation during replay attacks.

Table 8: Comparison of the original [11] and adapted version of CyPhERS.

Version	*TPR*	*FPR*	IR_{Scan}	IR_{HTTPS}	IR_{Spoof}	IR_{FCIA}	IR_{FDIA}	IR_{Replay}
Original	0.86	0.15	1/2	2/2	2/2	2/4	0/3	0/2
Adapted	0.98	0.02	2/2	2/2	2/2	4/4	3/3	2/2

6. Discussion

In this section, the results of applying CyPhERS for DER monitoring are discussed and put into a wider perspective.

6.1. Applicability of CyPhERS for monitoring of DERs and other power system applications

The results in Section 5 demonstrate that the proposed methodological adaptations and realization of an automated signature evaluation system enable application of CyPhERS for automated online DER monitoring. During all considered attacks, CyPhERS' Stage 1 provides event signatures which are automatically associated with the correct attack type, victim device, attacker location, and physical impact in Stage 2. In particular, the significant reduction of false positives and increased identification rate for cyber-physical attacks compared to the original version of CyPhERS (see Table 8) demonstrate the effectiveness of the proposed methodological adaptations, namely the 1) application of probabilistic models and detection rules, 2) differentiation of functionality- and behavior-describing physical target features, and 3) consideration of abstracting target features such as the on/off state of batteries. Given the complexity of the considered PV-battery system demonstration case, applying CyPhERS to other power system applications, including substations and energy communities, is considered possible. A potentially limiting factor for resource-constrained systems is the linear dependency between the number of models and target features. Thus, careful selection of monitored features is of high relevance for minimizing the computational burden of CyPhERS.

6.2. The role of ML in CyPhERS

In CyPhERS, ML is used to model target features and eventually provide the indicator for deciding whether an observation is normal or abnormal. The results in Section 5 demonstrate that ML allows to detect complex local anomalies which are only abnormal in a specific temporal or situational context (see, for example, Fig. 17). In case of some target features, similar detection results could be achieved with simpler methods. For example, the global anomaly in n_{ARP_d} during ARP spoofs (see Fig. 15 (a)) could be detected with a pre-defined static threshold. However, ML allows to generalize and automate modeling of target features and definition of detection rules. Thus, even in cases where simpler methods can achieve the same performance, ML is advantageous as it avoids manual effort, which is particularly relevant for larger numbers of target features. Furthermore, through regular retraining, the models automatically adapt to changes such as new consumer behavior.

6.3. *Uniqueness of event signatures*

For the sake of conciseness and readability, the number of target features (in particular network features) is limited in this work. Many other relevant features which are, for example, based on port numbers or MAC and IP addresses are neglected. Moreover, other information sources are fully excluded. These include human interactions with the system (e.g., maintenance activities), and system logs. Consequently, some of the event signatures may be explainable by other incidents as well. For example, the pattern of a FCIA may also result from the rare event of switching off inverters for maintenance. If models are informed about such activities, these events can be distinguished. Thus, for implementation outside an academic environment, all relevant target features should be taken into account, in order to guarantee uniqueness of the event signatures.

6.4. *Integration into a distributed attack detection system*

The steadily growing number of solar plants, battery storages, and electric vehicles makes coordinated malicious control of DER fleets an emerging opportunity for large-scale attacks against power systems. CyPhERS could provide the foundation of a bottom-up security architecture for power systems, which identifies such threats in a timely and reliable manner. Attack reports of multiple distributed CyPhERS systems could be aggregated and jointly evaluated by a cyber security incident response team (CSIRT). The CSIRT could then inform affected transmission or distribution system operators about cyber incidents in their area, including information on location, capacity and type of affected energy resources, to enable incident response such as isolation of affected DERs.

7. Conclusion

This work adapts and evaluates the **Cy**ber-**Ph**ysical **E**vent **R**easoning **S**ystem CyPhERS for automated online DER monitoring. CyPhERS is a two-stage process, where Stage 1 generates informative and interpretable signatures from an online evaluation of physical process and network traffic data, which are evaluated in Stage 2 to conclude on event root causes and physical impacts. Among the key strengths are the independence of historical event observations, and capability to provide information on cyber, physical and cyber-physical event types. To enable applicability of CyPhERS for DER monitoring, this work proposes and realizes several methodological adaptations for Stage 1, including 1) switching to probabilistic models and detection rules, 2) differentiating functional and behavioral target features, and 3) describing complex DER behavior via abstracting target features. Moreover, a rule-based system is formulated and implemented to automate signature evaluation in Stage 2. The applicability of the adapted version of CyPhERS for DER monitoring is evaluated on a dataset which describes several cyber and cyber-physical attack types targeting a real PV-battery system. The results demonstrate that the proposed methodological adaptations and rule-based signature evaluation system enable CyPhERS to automatically infer attack occurrence, type, victim devices, attacker location, and physical impact in all considered attack scenarios. The effectiveness of the methodological adaptations is particularly evident in significantly higher identification rates for cyber-physical attacks, and a reduction of the false positive rate from $FPR = 0.15$ to 0.02.

Acknowledgement

This work is partly funded by the Innovation Fund Denmark (IFD) under File No. 91363, and by the Helmholtz Association under the program ‘Energy System Design’.

References

[1] I. J. Perez-Arriaga, The transmission of the future: The impact of distributed energy resources on the network, IEEE Power and Energy Magazine 14 (4) (2016) 41–53. doi:10.1109/MPE.2016.2550398.

[2] M. Šarac, N. Pavlović, N. Bacanin, F. Al-Turjman, S. Adamović, Increasing privacy and security by integrating a blockchain secure interface into an iot device security gateway architecture, Energy Reports 7 (2021) 8075–8082. doi:https://doi.org/10.1016/j.egyr.2021.07.078.

[3] A. G. Eustis, The mirai botnet and the importance of iot device security, in: 16th International Conference on Information Technology-New Generations (ITNG 2019), Springer, 2019, pp. 85–89.

[4] S. Lakshminarayana, J. Ospina, C. Konstantinou, Load-altering attacks against power grids under covid-19 low-inertia conditions, IEEE Open Access Journal of Power and Energy 9 (2022) 226–240. doi:10.1109/OAJPE.2022.3155973.

[5] EnergiCERT, Cyber attacks against european energy & utility companies, `https://energicert.dk/wp-content/uploads/2022/09/Attacks-against-European-energy-and-utility-companies-2020-09-05-v3.pdf`, accessed: 2023-05-06.

[6] S. Huntley, Fog of war: how the ukraine conflict transformed the cyber threat landscape, `https://blog.google/threat-analysis-group/fog-of-war-how-the-ukraine-conflict-transformed-the-cyber-threat-landscape/`, accessed: 2023-05-06.

[7] Y. Li, J. Yan, Cybersecurity of smart inverters in the smart grid: A survey, IEEE Transactions on Power Electronics 38 (2) (2023) 2364–2383. doi:10.1109/TPEL.2022.3206239.

[8] N. Müller, C. Ziras, K. Heussen, Assessment of cyber-physical intrusion detection and classification for industrial control systems, in: 2022 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), 2022, pp. 432–438. doi:10.1109/SmartGridComm52983.2022.9961010.

[9] J. Ye, A. Giani, A. Elasser, S. K. Mazumder, C. Farnell, H. A. Mantooth, T. Kim, J. Liu, B. Chen, G.-S. Seo, W. Song, M. D. R. Greidanus, S. Sahoo, F. Blaabjerg, J. Zhang, L. Guo, B. Ahn, M. B. Shadmand, N. R. Gajanur, M. A. Abbaszada, A review of cyber–physical security for photovoltaic systems, IEEE Journal of Emerging and Selected Topics in Power Electronics 10 (4) (2022) 4879–4901. doi:10.1109/JESTPE.2021.3111728.

[10] N. D. Tuyen, N. S. Quan, V. B. Linh, V. Van Tuyen, G. Fujita, A comprehensive review of cybersecurity in inverter-based smart power system amid the boom of renewable energy, IEEE Access 10 (2022) 35846–35875. doi:10.1109/ACCESS.2022.3163551.

[11] N. Müller, K. Bao, J. Matthes, K. Heussen, Cyphers: A cyber-physical event reasoning system providing real-time situational awareness for attack and fault response (2023). arXiv:2305.16907.

[12] I. Zografopoulos, C. Konstantinou, Detection of malicious attacks in autonomous cyber-physical inverter-based microgrids, IEEE Transactions on Industrial Informatics 18 (9) (2022) 5815–5826. doi:10.1109/TII.2021.3132131.

[13] A. Y. Fard, M. Easley, G. T. Amariucai, M. B. Shadmand, H. Abu-Rub, Cybersecurity analytics using smart inverters in power distribution system: Proactive intrusion detection and corrective control framework, in: 2019 IEEE International Symposium on Technologies for Homeland Security (HST), 2019, pp. 1–6. doi:10.1109/HST47167.2019.9032978.

[14] Y. Li, P. Zhang, L. Zhang, B. Wang, Active synchronous detection of deception attacks in microgrid control systems, IEEE Transactions on Smart Grid 8 (1) (2017) 373–375. doi:10.1109/TSG.2016.2614884.

[15] S. Dey, M. Khanra, Cybersecurity of plug-in electric vehicles: Cyberattack detection during charging, IEEE Transactions on Industrial Electronics 68 (1) (2021) 478–487. doi:10.1109/TIE.2020.2965497.

[16] J. Zhang, L. Guo, J. Ye, Cyber-attack detection for photovoltaic farms based on power-electronics-enabled harmonic state space modeling, IEEE Transactions on Smart Grid 13 (5) (2022) 3929–3942. doi:10.1109/TSG.2021.3121009.

[17] S. Tan, J. M. Guerrero, P. Xie, R. Han, J. C. Vasquez, Brief survey on attack detection methods for cyber-physical systems, IEEE Systems Journal 14 (4) (2020) 5329–5339. doi:10.1109/JSYST.2020.2991258.

[18] L. Guo, J. Ye, B. Yang, Cyberattack detection for electric vehicles using physics-guided machine learning, IEEE Transactions on Transportation Electrification 7 (3) (2021) 2010–2022. doi:10.1109/TTE.2020.3044524.

[19] D. Said, M. Elloumi, L. Khoukhi, Cyber-attack on p2p energy transaction between connected electric vehicles: A false data injection detection based machine learning model, IEEE Access 10 (2022) 63640–63647. doi:10.1109/ACCESS.2022.3182689.

[20] A. Arsalan, L. Timilsina, B. Papari, G. Muriithi, G. Ozkan, P. Kumar, C. S. Edrington, Cyber attack detection and classification for integrated on-board electric vehicle chargers subject to stochastic charging coordination, Transportation Research Procedia 70 (2023) 44–51, 8th International Electric Vehicle Conference (EVC 2023). doi:https://doi.org/10.1016/j.trpro.2023.10.007.

[21] A. A. Khan, O. A. Beg, M. Alamaniotis, S. Ahmed, Intelligent anomaly identification in cyber-physical inverter-based systems, Electric Power Systems Research 193 (2021) 107024. doi:https://doi.org/10.1016/j.epsr.2021.107024.

[22] D. Mukherjee, A novel strategy for locational detection of false data injection attack, Sustainable Energy, Grids and Networks 31 (2022). doi:https://doi.org/10.1016/j.segan.2022.100702.

[23] Z. Warraich, W. Morsi, Early detection of cyber–physical attacks on fast charging stations using machine learning considering vehicle-to-grid operation in microgrids, Sustainable Energy, Grids and Networks 34 (2023). doi:https://doi.org/10.1016/j.segan.2023.101027.

[24] A. M. Kosek, O. Gehrke, Ensemble regression model-based anomaly detection for cyber-physical intrusion detection in smart grids, in: 2016 IEEE Electrical Power and Energy Conference (EPEC), 2016, pp. 1–7. doi:10.1109/EPEC.2016.7771704.

[25] M. Shaaban, U. Tariq, M. Ismail, N. A. Almadani, M. Mokhtar, Data-driven detection of electricity theft cyberattacks in pv generation, IEEE Systems Journal 16 (2) (2022) 3349–3359. doi:10.1109/JSYST.2021.3103272.

[26] T. Tabassum, O. Toker, M. R. Khalghani, Cyber–physical anomaly detection for inverter-based microgrid using autoencoder neural network, Applied Energy 355 (2024) 122283. doi:https://doi.org/10.1016/j.apenergy.2023.122283.

[27] C. B. Jones, A. R. Chavez, R. Darbali-Zamora, S. Hossain-McKenzie, Implementation of intrusion detection methods for distributed photovoltaic inverters at the grid-edge, in: 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2020, pp. 1–5. doi:10.1109/ISGT45199.2020.9087756.

[28] A. P. Kuruvila, I. Zografopoulos, K. Basu, C. Konstantinou, Hardware-assisted detection of firmware attacks in inverter-based cyberphysical microgrids, International Journal of Electrical Power & Energy Systems 132 (2021) 107150. doi:https://doi.org/10.1016/j.ijepes.2021.107150.

[29] C. B. Jones, A. Chavez, S. Hossain-McKenzie, N. Jacobs, A. Summers, B. Wright, Unsupervised online anomaly detection to identify cyber-attacks on internet connected photovoltaic system inverters, in: 2021 IEEE Power and Energy Conference at Illinois (PECI), 2021, pp. 1–7. doi:10.1109/PECI51586.2021.9435234.

[30] O. Avatefipour, A. S. Al-Sumaiti, A. M. El-Sherbeeny, E. M. Awwad, M. A. Elmeligy, M. A. Mohamed, H. Malik, An intelligent secured framework for cyberattack detection in electric vehicles' can bus using machine learning, IEEE Access 7 (2019) 127580–127592. doi:10.1109/ACCESS.2019.2937576.

[31] Y. Li, L. Zhang, Z. Lv, W. Wang, Detecting anomalies in intelligent vehicle charging and station power supply systems with multi-head attention models, IEEE Transactions on Intelligent Transportation Systems 22 (1) (2021) 555–564. doi:10.1109/TITS.2020.3018259.

[32] I. Zografopoulos, C. Konstantinou, N. D. Hatziargyriou, Distributed energy resources cybersecurity outlook: Vulnerabilities, attacks, impacts, and mitigations, arXiv preprint arXiv:2205.11171 (2022).

[33] V. K. Singh, M. Govindarasu, A cyber-physical anomaly detection for wide-area protection using machine learning, IEEE Transactions on Smart Grid 12 (4) (2021) 3514–3526. doi:10.1109/TSG.2021.3066316.

[34] A. Chavez, C. Lai, N. Jacobs, S. Hossain-McKenzie, C. B. Jones, J. Johnson, A. Summers, Hybrid intrusion detection system design for distributed energy resource systems, in: 2019 IEEE CyberPELS (CyberPELS), 2019, pp. 1–6. doi:10.1109/CyberPELS.2019.8925064.

[35] A. Sahu, Z. Mao, P. Wlazlo, H. Huang, K. Davis, A. Goulart, S. Zonouz, Multi-source multi-domain data fusion for cyberattack detection in power systems, IEEE Access 9 (2021) 119118–119138. doi:10.1109/ACCESS.2021.3106873.

[36] A. A. Cook, G. Mısırlı, Z. Fan, Anomaly detection for iot time-series data: A survey, IEEE Internet of Things Journal 7 (7) (2020) 6481–6494. doi:10.1109/JIOT.2019.2958185.

[37] M. K. Hasan, A. A. Habib, Z. Shukur, F. Ibrahim, S. Islam, M. A. Razzaque, Review on cyber-physical and cyber-security system in smart grid: Standards, protocols, constraints, and recommendations, Journal of Network and Computer Applications 209 (2023) 103540. doi:https://doi.org/10.1016/j.jnca.2022.103540.

[38] F. Li, X. Yan, Y. Xie, Z. Sang, X. Yuan, A review of cyber-attack methods in cyber-physical power system, in: 2019 IEEE 8th International Conference on Advanced Power System Automation and Protection (APAP), 2019, pp. 1335–1339. doi:10.1109/APAP47170.2019.9225126.

[39] B. E. Strom, A. Applebaum, D. P. Miller, K. C. Nickels, A. G. Pennington, C. B. Thomas, Mitre att&ck: Design and philosophy, in: Technical report, The MITRE Corporation, 2018.

[40] S. Lakshminarayana, S. Adhikari, C. Maple, Analysis of iot-based load altering attacks against power grids using the theory of second-order dynamical systems, IEEE Transactions on Smart Grid 12 (5) (2021) 4415–4425. doi:10.1109/TSG.2021.3070313.

[41] Q. Wang, Y. Ma, K. Zhao, Y. Tian, A comprehensive survey of loss functions in machine learning, Annals of Data Science (2020) 1–26.

[42] J. H. Friedman, Greedy function approximation: A gradient boosting machine, The Annals of Statistics 29 (5) (2001) 1189–1232.
URL `http://www.jstor.org/stable/2699986`

[43] C. S. Bojer, J. P. Meldgaard, Kaggle forecasting competitions: An overlooked learning opportunity, International Journal of Forecasting 37 (2) (2021) 587–603. doi:https://doi.org/10.1016/j.ijforecast.2020.07.007.