Model-Free Detection of Cyberattacks on Voltage Control in Distribution Grids

Mohammed S. Kemal[*], Wissam Aoudi[†], Rasmus L. Olsen[*], Magnus Almgren[†], Hans-Peter Schwefel[*]

[*]Aalborg University, Aalborg, Denmark

{seifu, rlo, hps}@es.aau.dk

[†]Chalmers University of Technology, Gothenburg, Sweden

{wissam.aoudi, magnus.almgren}@chalmers.se

Abstract—Incorporating information and communication technology in the operation of the electricity grid is undoubtedly contributing to a more cost-efficient, controllable, and flexible power grid. Although this technology is promoting flexibility and convenience, its integration with the electricity grid is rendering this critical infrastructure inherently vulnerable to cyberattacks that have potential to cause large-scale and far-reaching damage. In light of the growing need for a resilient smart grid, developing suitable security mechanisms has become a pressing matter. In this work, we investigate the effectiveness of a model-free state-of-the-art attack-detection method recently proposed by the cybersecurity community in detecting common types of cyberattacks on voltage control in distribution grids. Experimental results show that, by monitoring raw controller and smart-meter data in real time, it is possible to detect denial of service, replay, and integrity attacks, thus contributing to a resilient and more secure grid.

Index Terms—Low-Voltage Grid, Cyberattack, Model-Free Detection, Smart Grid, PASAD

I. Introduction

Due to limitations, costs, and growing concerns over environmental impact of the electricity grid, transitioning into the envisioned cost-effective, more environment-friendly, highly manageable and controllable *smart grid* has become increasingly pressing over the past few years. Advances in information and communication technology—the driving force behind this transition—are paving the way for a more flexible distribution grid capable of resolving the limitations of the current electricity grid and optimizing the integration of renewable energy sources, such as wind and solar power. The other side of the coin, however, is that the integration of communication technologies makes the smart grid susceptible to cyberattacks capable of causing serious damage to the electricity infrastructure.

The successful operation of smart grid services, such as monitoring and control of low-voltage distribution grids (LV-grids), demand management, energy theft detection and load forecasting [1], relies heavily on fine-grained smart meter readings. The transmission of such sensitive data over insecure communication links, however, goes beyond privacy issues and opens doors to malicious actors to compromise the grid operation via cyberattacks that could cause, for instance, a massive operational failure of energy assets [2].

In light of the expanding threat surface in energy networks, the ability to detect cyberattacks before they cross from the cyber realm to the physical world is growingly needed. In this paper, we investigate the effectiveness of PASAD, a recently proposed model-free technique for detecting attacks on industrial control systems, in detecting various common types of cyberattacks on LV-grids. PASAD captures the dynamics of voltage-control loops during a training phase through sophisticated analysis of time series of controller and smart-meter data, then detects deviations from the normal behavior through real-time data processing. The motivation behind using a model-free detection approach in current LV grids is twofold. First, the data required for modelling current distribution grids is scarce and often inaccurate. This applies to both household consumption measurements and grid topology data such as line lengths, cable types and cable parameters. Hence, model-based techniques are difficult to apply as they require correlating measurements with a model of the LV grid. Second, a model-free approach is inherently agnostic to the controller scenario and can thus be used for different kinds of control, independently of the underlying LV grid.

Control of photo-voltaic (PV) and small-scale wind turbines in electricity distribution grids is growingly adopted to address challenges arising from the recent proliferation of distributed generation units. The voltage-control scenario under study in this paper targets reduction of over- and under-voltages by adjusting reactive power generation of selected LV grid assets. The proposed approach is validated through a series of experiments using a low-voltage grid model based on a single-phase representation of a realistic LV grid in Denmark. Experimental results using various representative low-voltage control scenarios demonstrate a promising capability of detecting subtle attack-induced changes in the system behavior.

The remainder of this paper is laid out as follows: In Section II, we present background material and discuss related work. Section III describes the system architecture and Section IV introduces the attack scenarios and illustrates the attacker model. The attack-detection methodology is presented in Section V and the proposed approach is evaluated in Section VI. Finally, we conclude this work in Section VII.

II. Background

A. Low-Voltage Distribution Grids

The operation of future LV-grids typically involves the control and coordination of distributed renewable generation
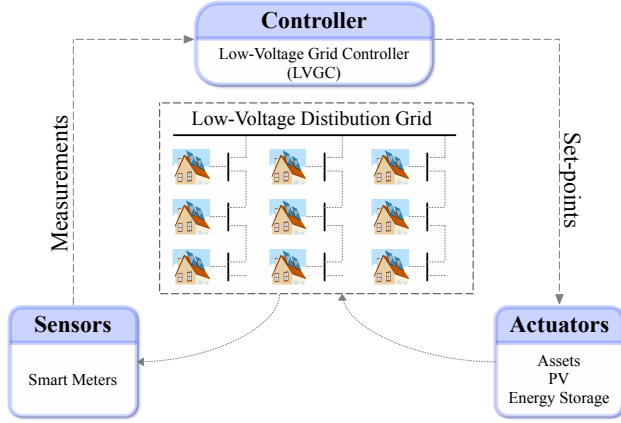
171

Fig. 1: A conceptual control loop for maintaining low-voltage grids within the operational bounds.

Fig. 2: Layout and architecture of the reference LV-grid used in this work. The nodes in green boxes are controllable assets [6].

and consumption units, collectively referred to as *controllable assets*. In particular, wind turbines and PV systems scattered throughout distribution networks exchange system states with the controller [3,4]. Figure 1 outlines a conceptual control loop as well as a high-level description of the information flow in a typical voltage-control scenario in an LV-grid. Sensors measuring the grid state communicate their measurements via smart meters (SMs) to a voltage controller over a communication network that is necessarily fast and reliable to optimize the use of controllable assets [5,6]. Based on a predefined control objective, the controller then utilizes the received grid state information to make decisions on how actuators in the grid should operate; for example, calculating new set-points for PV systems or energy storage.

The voltage control implementation used in this work is event-driven. In an event-driven control setting, the control loop is triggered upon violation of certain control criteria. In the event of such a breach, the controller starts executing periodically until the grid satisfies the control criteria. In an attempt to minimize the potential economic implications of prolonged continuous control, the event-driven controller remains idle for as long as possible.

B. Related Work

LV grids are witnessing a rapidly increasing integration of distributed inverter-based generation. Although the distributed generation units may contribute to voltage problems, such as over-voltages, harmonics, dips, and swells, they could also be leveraged to solve the very same problems. In an increasingly common control situation, inverters participate in intelligent grid controls to solve the voltage problems [7,8].

Ma et al. [9] discuss measurement falsification scenarios, wherein an adversary corrupts voltage measurements received by so-called voltage droop controllers, and assess the impact of such attacks on system stability and voltage magnitude using analytical control-theoretic methods. Using a rather simple grid model, the authors focus solely on the effects of such attacks on controllers and do not investigate attack detection.
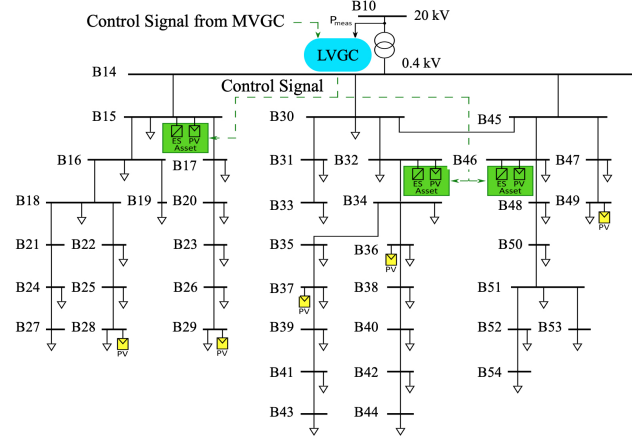
In [10,11], so-called cyber-secure modeling frameworks for the power grid and the communication networks are discussed. Kundur et al. [11] consider general smart-grid scenarios, whereas Giacomoni et al. [10] explicitly propose an intelligent distributed secure control architecture for distribution systems to provide greater adaptive protection through proactive reconfiguration and rapid response to disturbances. In both works, however, the impact and detection of cyberattacks are not addressed.

Isozaki et al. [12] propose a detection algorithm for centralized voltage regulation, whereby voltage measurements from sectioning switches equipped with sensors feeding to a centralized controller are monitored to detect attacks that are performed at the controller level only.

This paper studies the consequences of cyberattacks on voltage control in LV distribution grids and proposes a model-free approach to detecting such attacks on both the controller's and actuators' side.

III. SYSTEM DESCRIPTION

The layout and architecture of the reference LV-grid used to validate our proposed approach are presented in Fig. 2. In the setup shown therein, local measurements of voltage and total power on the LV bus bar are accessed by a low-voltage grid controller (LVGC) assumed to be located in a secondary substation. The three nodes highlighted in green represent the controllable assets used in our evaluation, which consist of PV systems equipped with energy storage capabilities. Sensors mounted on the controllable assets measure the local voltage for the controllable assets, as well as the minimum and maximum active and reactive power, and communicate with smart meters to notify the controller in the event of a threshold violation. To keep the voltages within operational bounds, the controller communicates set-points to the controllable assets. The reference grid is simulated using the Matlab-based tool DiSC, which is an open-source simulation framework

172

originally developed to verify voltage-control approaches in European power distribution systems [13]. In order to reproduce realistic dynamics in the reference grid, in addition to simulating household consumption patterns for each node in the grid, some of the nodes are equipped with PV systems and storage elements, thereby adding more variability to the voltage behaviour. This variability is demonstrated in Fig. 3, which shows a sample voltage behavior of two nodes with and without PV systems.

For the purpose of this work, we use a generalized event-driven voltage-control strategy wherein the controller starts executing whenever voltage measurements at the asset nodes violate a prespecified threshold. The LVGC controls the behavior of the reactive power by communicating control signals (set-points) to the controllable assets. Upon a threshold violation, the voltage controller runs every 2 minutes and sends set-points to the controllable assets in the grid until the voltage recovers the normal level. The control for each individual asset is a droop control changing solely the reactive power according to the equation $Q_{ref}(t) = G_D \times (1 - |VOutAssets(t)|/V_{base})$, where $VOutAssets(t)$ is the measured asset's voltage at time t, V_{base} is the nominal voltage, and G_D is the droop gain obtained by manual tuning of the controller. Although it processes local information for control, the use of a centralized controller is motivated by the possible coordination of assets in reaching a global objective [8].

IV. ADVERSARY MODEL AND ATTACK SCENARIOS

Previous studies have shown that many existing smart meters lack the necessary means of ensuring data integrity and authenticity [14,15]. Although the damage inflicted by compromising a few smart meters and causing voltage fluctuations may be limited to household equipment, an attacker taking control of a sufficiently large subset of smart meters may be able to destabilize the entire infrastructure, potentially leading to nothing less than a blackout.

The adversary model assumed in this work is presented in Fig. 4. The voltage-control loop is exploitable by an adversary whose objective is to destabilize the distribution grid. Due to the closed-loop structure, small malicious modifications to the control signals can be iteratively amplified by the control loop, causing an increased violation of voltage thresholds, drainage of energy storage, and increased system operating cost.
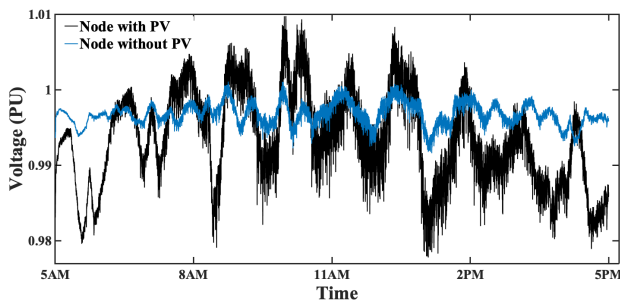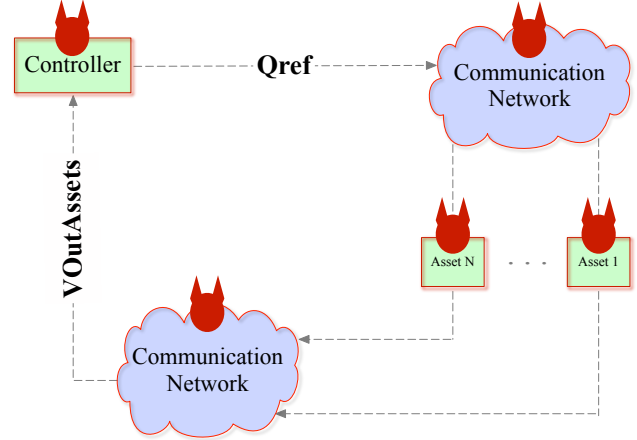

Fig. 3: Voltage dynamics with and without PV systems.

Fig. 4: The adversary model.

As shown in Fig. 4, on one hand, the adversary can manipulate the control signals Qref by either compromising the communication link between the controller and the controllable assets or by directly compromising the controller, in which case the assets would execute malicious signals sent by the adversary. On the other hand, the adversary may compromise smart meters and manipulate the VOutAssets readings so that the controller reacts erroneously.

We consider three common attack types, namely, denial of service, replay, and integrity attacks. Following is a description of the attack scenarios considered in this work.

Denial of Service Attack. In an event-based voltage-control setting, the integration of computing systems, communication networks, and physical electric power systems gives rise to a multi-dimensional and heterogeneous complex environment with real-time sensing, dynamic control, and information services. A Denial of Service (DoS) attack is a resource-exhausting attack that effectively suspends the control of the system in an attempt to bring it to an unsafe state. A DoS attack is launched by attackers whose aim is to cause lack of service availability, e.g., a power outage affecting customers and distribution system operators alike. For a distribution grid with a networked control system, DoS attacks can take many forms [16]. For instance, the attacker can flood the communication network with useless requests to exhaust the network resources and thus suspend the exchange of messages carrying control signals. Alternatively, attackers could compromise the controller entirely and block access to the communication channel or completely switch off the controller.

Replay Attack. In a replay attack, the adversary replays previously recorded traffic in an attempt to fool the controller. As discussed in [17], a successful replay attack initially involves collecting and passively recording sequences of data by manipulating the controller, the communication network, or the smart-meter measurements. The previously recorded data sequences are subsequently replayed onto the network during a desired time interval. A successful replay attack does not require prior knowledge of the system components.

173

Integrity Attack. We consider integrity attacks on the control signals sent by the controller to the controllable assets. In an integrity attack, the Qref control signal is maliciously manipulated by the adversary so that the set-point in the current control loop received by the asset differs from the true set-point sent by the controller. Unlike replay attacks, integrity attacks require extensive domain-knowledge of the components and operation of the target system. Specifically, conducting an integrity attack requires the adversary to be capable of modifying the controller data at the controller, during transmission, or at the smart meters. Integrity attacks on control signals can be performed in different ways by exploiting known vulnerabilities [15]. For instance, by compromising intermediate nodes in the communication network of the power grid (e.g., routers), an attacker can intercept and forge network packets carrying Qref signals so that they contain maliciously altered set-points.

V. ATTACK DETECTION METHODOLOGY

To detect the simulated attacks on the LV-grid described in Section III, we apply a Process-Aware Stealthy-Attack Detection mechanism (PASAD) that has recently been proposed by Aoudi et al. [18] to detect attacks on industrial control systems. The method takes as input a time series of process measurements and raises an alarm whenever a change in system dynamics is suspected.

PASAD works in two phases: an offline training phase and an online detection phase. In the training phase, the normal behavior of the underlying dynamical system is represented mathematically in a low-dimensional *signal subspace* by means of spectral decomposition of a special matrix derived from the time-series data. Afterwards, during the detection phase, the most recent process measurements are compared to the normal behavior established in the training phase to determine whether or not a structural attack-indicating change in behavior is taking place. This is done by computing a *departure score* for every new measurement to determine the extent to which current readings conform to the estimated dynamics. Finally, an alarm is raised whenever the computed score crosses a certain threshold determined during a validation period.

We run two concurrent instances of PASAD to monitor the two different time series of process data that exist in the control loop (see Fig. 4): smart meter voltage measurements VOutAssets and controller set-points Qref. For each of the two instances, an initial subseries of the measurements time series is used to construct a so-called *trajectory matrix* $\mathbf{X}$. Then, a subset of the eigenvalues of the covariance matrix $\mathbf{XX^T}$, obtained from the *singular value decomposition* of $\mathbf{X}$, is selected to form a basis for the low-dimensional signal subspace. When vectors constructed out of the time series of measurements are projected onto the signal subspace, the following phenomenon occurs: under normal operating conditions, the projected vectors occupy a bounded region and thereby form a cluster, whereas under attack conditions, the vectors *depart* from the cluster. To measure this departure, a departure score is computed for the most recent test vector to determine the distance from the cluster.

A succinct depiction of the workings of PASAD is presented in Fig. 5. The time series shown in the left plot is an artificial square wave with added white Gaussian noise. During the training phase, PASAD processes an initial part of the time series to learn about the underlying signal by identifying a mapping from the input space to a so-called signal subspace as shown in the right plot. The mapping is an orthogonal projection that transforms the time-series measurements into vectors in a low-dimensional vector space in which consecutive training vectors follow a pattern, thereby establishing a baseline of normal system behavior. During the testing phase, PASAD continuously checks if the most recent test vectors conform to the pattern. In the event of a structural change in the time series, as shown in the figure, the test vectors break out of the pattern indicating an anomalous behavior.

For a comprehensive treatment of the underlying theory and parameters setting, the reader may refer to [18].

VI. EVALUATION

In this section, we introduce the experimental setup, followed by a description of the experiments corresponding to the three attack scenarios introduced in Section IV.

For all the experiments, the upper subplot displays the time series of process measurements monitored by PASAD comprising five days of operation. The initial subseries highlighted in blue was used for training and estimating the signal subspace. In all cases, the attack takes place on the fifth day at noon and lasts for two hours. The measurements during the attack interval are highlighted in red. The lower subplot shows the departure scores computed iteratively by PASAD for every new measurement, together with the alarm threshold. The threshold was determined by running PASAD for a validation period of 24 hours and then selecting the maximum value attained plus a relatively small constant. For each attack scenario, two experiments were conducted: one where Qref was monitored and another where VOutAssets was monitored. For the sake of producing difficult attack cases, the simulation was performed during summer days, where the attacks occured at lunchtime, which is the time when the controller is most active due to the unpredictable stochastic behavior of the PV systems induced by solar irradiance.

A. Experimental Setup

The LV-grid simulation model used in our experiments is an extrapolation of an actual LV-grid in northern Denmark. The simulated grid comprises 37 housing units with integrated heat pumps (see Fig. 2) among which 8 units have uncontrollable PV systems and 3 units have controllable PV systems combined with energy storage (battery). The simulated household consumption patterns are based on real consumption models. The behavior of the PV systems is generated by models incorporating solar irradiance, geographical location, cloud cover, and time of the day [6,13,19]. For the experiments in this work, the threshold bounds are $1\pm0.013pu$ for identifying voltage
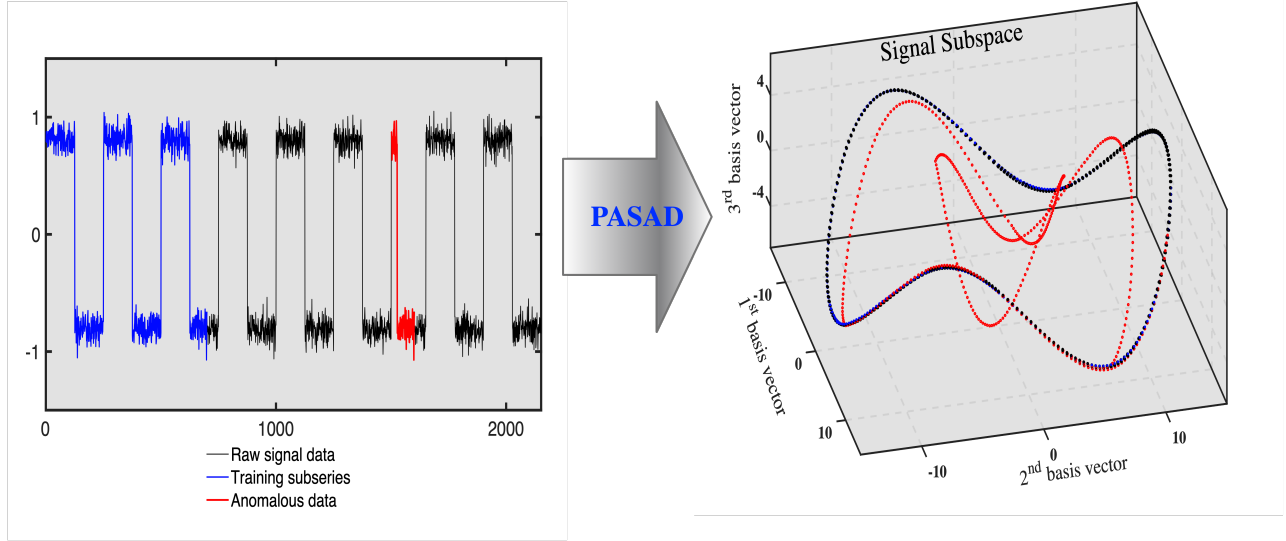
174

Fig. 5: A visual demonstration of how PASAD detects anomalies in time series by transforming the signal data into a geometric space wherein the anomalous behavior (highlighted in red) is easier to detect.

violations and $1 \pm 0.01 pu$ for activating the controller. The controller activation bounds are narrower to give the controller time in advance of the voltage violation. Furthermore, these bounds were chosen to be relatively small in order to activate the controller more frequently. Regarding the droop control, the single-phase nominal voltage V_{base} is set to $400V$, and the droop gain (G_D) is chosen to be 2×10^5 VAr/V in the experiments based on manual tuning of the controller. The parameters of the LV-grid, asset models, and the generalized event-driven controller are summarized in Table I.

B. The DoS Attack Experiment

To perform the DoS attack, we assume that the attacker compromises the control center and switches off the controller for two hours, which, according to the simulation model used in this work, causes the control signals (Qref) to go to zero during the attack. The detection results of this attack are displayed in Fig. 6a and Fig. 6b. As shown in the

TABLE I: Reference Grid Parameters

Parameter	Value
Number of electrical nodes	49
Number of buses	42
LV base voltage	400V
PV max rated power	6kW
PV efficiency	20%
PV area	28m^2
Energy storage capacity	65kWh
Energy storage rated power output	10kW
Summer simulation day	June 9
Winter simulation day	February 2
Geographical latitude	56,889^o

figures, the attack was detected by PASAD in both Qref and VOutAssets.

C. The Replay Attack Experiment

For the replay attack scenario, the attacker is assumed to have the means to passively record Qref control signals transmitted by the controller towards the sensor at the asset

(a) DoS attack on set-points   (b) DoS attack on sensors

(c) Replay attack on set-points   (d) Replay attack on sensors

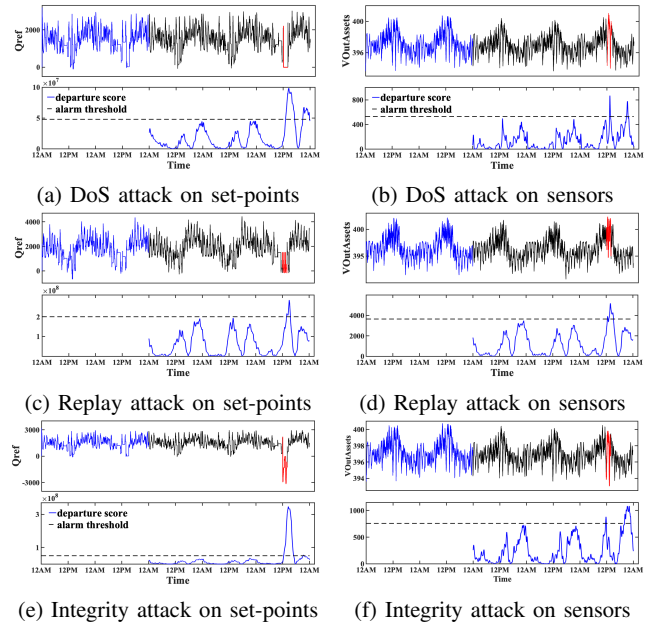(e) Integrity attack on set-points   (f) Integrity attack on sensors

Fig. 6: Detection of DoS, replay, and integrity attacks on both Qref set-points and VOutAssets measurements.

175

side. This can be achieved by compromising the wireless communication channel between the sensor and the controller or by gaining full access to the sensor interface. The recording occurs from 10AM to 11AM. The recorded traffic is subsequently replayed by the attacker during lunchtime from 12PM to 2PM. Both the recording and replaying occured on the fifth day of operation. The results of this experiment are displayed in Fig. 6c (for `Qref`) and Fig. 6d (for `VOutAssets`). As can be seen in the figures, PASAD successfully detects the replay attack at both ends of the control loop.

D. The Integrity Attack Experiment

To simulate an integrity attack on LV-grids, we assume that the attacker, by compromising the communication link, forges the set-points sent by the controller to the controllable assets. Specifically, the attacker alters the control signals in such a way that they perform the opposite function, i.e., injecting instead of consuming reactive power or consuming instead of injecting reactive power, where the latter is the case in this experiment. As in the previous scenarios, the attack was detected in both the control signals and the sensor measurements as shown in Fig. 6e and Fig. 6f respectively.

E. Discussion

As experimental results show, cyberattacks on LV-grids can be detected using a lightweight data-driven approach that obviates the need for building complex models and predicting future grid states. Due to the closed-loop mechanism, attacks on the controller manifest structural changes in the smart-meter readings. In all test cases, PASAD managed to detect the attacks at both ends of the controller. Note that in some cases, the controller does not recover the normal behavior completely, which explains the second spike in the departure scores after the attack onset. Also, it should be pointed out that in the integrity attack scenario, although the controller exhibits an implausible behavior (Fig. 6e), it is possible for a strategic adversary to spoof the control signals while carrying out the attack. However, as demonstrated in Fig. 6f, the attack can still be detected at the other end of the control loop. Although our analysis and results were based a single grid scenario in Denmark, the grid model used in our experiments features a high integration of renewable units and incorporates real household consumption measurements as well as grid-topology data such as line lengths, cable types and cable parameters, thus contributing to richer dynamics. It is therefore likely that the proposed approach is applicable to most grid models that involve less integration of renewable units.

VII. Conclusion

The current electricity grid appears to be taking steady steps towards the more efficient, modernized smart grid, with substantial integration of information and communication technologies seeming inevitable in the process. Being a critical infrastructure, securing the grid against cyberattacks proves necessary. In particular, a likely consequence of cyberattacks on the LV-grid is the contamination of measurements collected from compromised nodes, which in turn leads to bad control decisions by the controller. If a proper attack-detection mechanism is in place, compromised assets can be excluded by the controller when making control decisions. In this paper, we proposed a systematic approach to detecting cyberattacks on voltage control in LV-grids by monitoring time series of process data. Experimental results show that various common types of cyberattacks (DoS, replay, integrity) on LV-grids can be successfully detected using a model-free data-driven approach that does not require building complex models of the underlying system. As we have shown, monitoring control signals and voltage measurements both at the controller's side and the assets' side is a particularly effective attack-detection architecture. Future work is planned to consider more sophisticated stealthy integrity attacks, and investigate the detection capabilities in other control scenarios, such as power balancing and medium voltage-control.

References

[1] F. Skopik, "Security is not Enough! On Privacy Challenges in Smart Grids," *International Journal on Smart Grid and Clean Energy*, 2012.

[2] MARSH, "Could Energy Industry Dynamics Be Creating an Impending Cyber Storm?" https://urlzs.com/G2F42, 2018.

[3] C.-H. Lo and N. Ansari, "Decentralized Controls and Communications for Autonomous Distribution Networks in Smart Grid," *IEEE Transactions on Smart Grid*, 2013.

[4] A. Bidram and A. Davoudi, "Hierarchical Structure of Microgrids Control System," *IEEE Transactions on Smart Grid*, 2012.

[5] R. Pedersen, M. Findrik, C. Sloth, and H.-P. Schwefel, "Network Condition Based Adaptive Control and its Application to Power Balancing in Electrical Grids," *Sustainable Energy, Grids and Networks*, 2017.

[6] M. Kemal, L. Petersen, F. Iov, and R. L. Olsen, "A Real-Time Open Access Platform towards Proof of Concept for Smart Grid Applications," *Journal of Communication, Navigation, Sensing and Services (CONASENSE)*, 2017.

[7] P. Aristidou, G. Valverde, and T. Van Cutsem, "Contribution of Distribution Network Control to Voltage Stability: A Case Study," *IEEE Transactions on Smart Grid*, 2017.

[8] T. le Fevre Kristensen, R. L. Olsen, J. G. Rasmussen, and H.-P. Schwefel, "Information Access for Event-Driven Smart Grid Controllers," *Sustainable Energy, Grids and Networks*, 2017.

[9] M. Ma, A. M. Teixeira, J. Van Den Berg, and P. Palensky, "Voltage Control in Distributed Generation under Measurement Falsification Attacks," *IFAC-PapersOnLine*, 2017.

[10] A. M. Giacomoni, S. M. Amin, and B. F. Wollenberg, "A Control and Communications Architecture for a Secure and Reconfigurable Power Distribution System: An Analysis and Case Study," in *18th IFAC World Congress, Milano, Italy*, 2011.

[11] D. Kundur, X. Feng, S. Mashayekh, S. Liu, T. Zourntos, and K. L. Butler-Purry, "Towards Modelling the Impact of Cyber Attacks on a Smart Grid," *International Journal of Security and Networks*, 2011.

[12] Y. Isozaki, S. Yoshizawa, Y. Fujimoto, H. Ishii, I. Ono, T. Onoda, and Y. Hayashi, "Detection of Cyber Attacks against Voltage Control in Distribution Power Grids with PVs," *IEEE Transactions on Smart Grid*, 2016.

[13] R. Pedersen, C. Sloth, G. B. Andresen, and R. Wisniewski, "DiSC: A Simulation Framework for Distribution System Voltage Control," in *2015 European Control Conference (ECC)*, July 2015.

[14] S. McLaughlin, D. Podkuiko, and P. McDaniel, "Energy theft in the advanced metering infrastructure," in *International Workshop on Critical Information Infrastructures Security*. Springer, 2009.

[15] I. Rouf, H. Mustafa, M. Xu, W. Xu, R. Miller, and M. Gruteser, "Neighborhood Watch: Security and Privacy Analysis of Automatic Meter Reading Systems," in *Proceedings of the 2012 ACM conference on Computer and Communications Security*. ACM, 2012.

[16] P. Srikantha and D. Kundur, "Denial of Service Attacks and Mitigation for Stability in Cyber-Enabled Power Grid," in *Innovative Smart Grid Technologies Conference (ISGT), Power & Energy Society*. IEEE, 2015.

[17] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson, "Attack Models and Scenarios for Networked Control Systems," in *Proceedings of the 1st International Conference on High Confidence Networked Systems*. ACM, 2012.

[18] W. Aoudi, M. Iturbe, and M. Almgren, "Truth Will Out: Departure-Based Process-Level Detection of Stealthy Attacks on Control Systems," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2018.

[19] R. Pedersen, C. Sloth, and R. Wisniewski, "Active Power Management in Power Distribution Grids: Disturbance Modeling and Rejection." IEEE, 2016.