IEEE Open Access Journal of Power and Energy

Received 22 November 2023; revised 18 July 2024; accepted 29 September 2024.
Date of publication 31 October 2024; date of current version 13 November 2024.

Digital Object Identifier 10.1109/OAJPE.2024.3489477

Model-Based Detection of Coordinated Attacks (DCA) in Distribution Systems

NITASHA SAHANI (Member, IEEE), AND CHEN-CHING LIU (Life Fellow, IEEE)

Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061 USA

CORRESPONDING AUTHOR: N. SAHANI (nitashas@vt.edu)

This work was supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) through the Solar Energy Technologies Office under Award DE-EE0009339.

ABSTRACT The fast-paced growth in digitization of smart grid components enhances system observability and remote-control capabilities through efficient communication. However, enhanced connectivity results in heightened system vulnerability towards cybersecurity risks in the cyber-physical power system. Coordinated cyber-attacks (CCA), when undetected, lead to system-wide impact in terms of large disturbances or widespread outages. Detecting CCA in the cyber layer is critical to thwart cyber-attacks in real-time before the attack impacts the physical system. The challenge of locating CCA stems from the complex grid dynamics, making it difficult to distinguish between normal operational variations and cyber-attack impact. CCA often employs multiple attack vectors targeting geographically distributed components, further complicating CCA identification. Existing research in intrusion detection is primarily focused on the transmission network and limited to detecting individual attacks. In this paper, a novel proactive DCA strategy is proposed for early detection of CCA by establishing correlations among distinct attack events through model-based reinforcement learning that utilizes abductive reasoning to conclude the attacker goal. The solution includes understanding the system model, learning the system dynamics, and correlating individual cyber-attacks to extract the attacker's objective. The developed learning algorithm identifies the most probable attack path to reach the attacker's objective by predicting the next attack steps. A DNP3-based cyber-physical co-simulation testbed is developed to test the proposed algorithm using the IEEE 13-node test feeder.

INDEX TERMS Abductive reasoning, coordinated attacks, cyberattacks, cyber-physical systems, cybersecurity, distribution systems, intrusion detection, model-based reinforcement learning, smart grid.

I. INTRODUCTION

THE traditional power grid has evolved into an increasingly complex smart grid with the integration of a physical grid and a coupled cyber system. This is facilitated by a highly interconnected communication system to enhance remote monitoring and control of bi-directional flow of power and information. With enhanced observability, the system also becomes vulnerable due to an expansion of the cyber-attack surface. The impact of a cyber-attack on the distribution system can range from minor disturbances in node voltage levels to cascading events leading to power outages.

In recent times, cyber-attacks on the critical infrastructure have become more frequent. Stuxnet (2010) [1], Shamoon malware cyberattack (2012) [2], [3], and colonial pipeline ransomware attack (2021) [4] are prominent examples. The impact of the cyber-attack on the Ukrainian power grid (2015) affected several substations and led to widespread outages [5]. The impact of this attack confirmed that coordinated cyber-attacks can severely damage the physical grid. Cybersecurity concerns have been addressed through research in the areas to detect cyber-attacks, prevent the impact of an attack, and restore the system following the attack.

Traditional measures for cybersecurity of the Supervisory Control and Data Acquisition (SCADA) systems are focused on information security such as firewalls, abnormal data detection and encryption methods to ensure data security. With increased adoption of distributed devices, numerous at the grid edge, intrusion detection systems (IDS) are deployed at centralized distribution management system and/or at distributed node levels. These IDS can be broadly categorized to detect attacks in *i*) operational technologies (e.g., state estimation attacks) [6], and *ii)* communication/ cyber system

(e.g., communication layer congestion, data packets manipulation) [7], [8].

IDS for distribution systems is an evolving research area [9], [10]. Apart from the usual cyber network attacks (e.g., sniffing, reconnaissance, and Denial of Service), power distribution systems are prone to data integrity attacks such as false data injection (FDI), packet modification, and replay attacks [11]. Distribution systems operate at low-to-medium voltage levels. Compared to individual cyber-attacks at the distribution node level, multiple simultaneous attacks (attacking multiple power components at the same time) [12] or sequential attack schemes (attacking multiple power components in a time sequence) [13] have the potential to cause significant damage to the system. There has been considerable detection effort in modeling and detecting stealth deception attacks at the distribution systems with high distributed energy resources penetration [14].

Recent studies primarily focus on detecting individual attacks which although provides an initial screening step, cannot capture the underlying relationship between multiple attacks in the system. A Coordinated Cyber-Attack (CCA) is characterized by strategically curated multiple attack events that can be effectively correlated, aiming for a higher attack goal. It can be multiple attacks attempted at one component or a strategic series of attacks on different components for a specific goal. Hence, CCA has 3 main components: i) multiple attack events at one/ multiple locations; ii) correlation between attacks; and iii) ultimate attack goal. Lack of a systematic methodology to identify the attack correlations will fail to defend the system against multi-step coordinated attacks. It is, therefore, critical to develop an approach for intrusion prevention by extracting attack patterns to trace the possible attack objectives based on correlations between the generated cyber system alarms and corresponding physical system behavior.

A coordinated attack detection system is developed at the substation level by use of correlation indices to capture attack patterns in the form of abnormal behavior, substation criticality, and geographical location [15]. This method utilizes three specific hard-wired relation features which will not be able to detect attack sequences if it does not follow these relations. Reference [16] proposed a method to mine attack patterns using a decision tree based on temporal-topological correlation between cyber-attack sequences and physical attack events. However, the offline analysis will not be able to detect and prevent coordinated attacks in real-time. A multi-agent intrusion prevention system is formulated in [17] against CCA in the distribution system. Attack correlation is measured based on intrusion potential, criticality, and vulnerability indices using a decentralized decision-making consensus algorithm.

Learning-based algorithms have been studied to detect cyber-attacks in a smart grid [18]. Integration of physical measurements and cyber data is modeled as a data link in [19] for detecting CCA in a cyber-power system (CPS), using gradient boosting decision trees with principal component analysis. Deep learning is used to detect coordinated FDI attacks on distributed generation [20]; however, supervised learning needs a high volume of relevant system data for extensive training, which is usually lacking in the power domain. Due to the limitation of data availability, multiple studies have used reinforcement learning (RL) for CCA detection. A multi-agent system correlates attacks and predicts the CCA target, after which Q-learning is used to mitigate the attack [21]. Similarly, Q-learning is modified to design an attacker-defender zero-sum game [22] to detect the multistage attack sequence. A coordinated cyber-physical topology attack is identified using deep Q-learning [23] to provide the vulnerability measure of critical transmission lines. The cyber-resiliency requirements for the DER-based smart grid along with detailed intrusion detection methodologies are compared in [24] while recognizing the possible cyber enhancement trajectories.

Although model-free methods have been used for intrusion detection, the proposed model-based reinforcement learning (MBRL) for attack detection is a new concept for the power grid. There have been studies that use MBRL for power system control applications [25], [26]. However, MBRL for cybersecurity has not been fully explored.

Existing methods and algorithms for CCA detection have significant contributions. However, there are some limitations:

1) Present work on CCA, primarily focused on transmission networks, is not directly scalable to distribution systems due to its distributed nature and low computational capability of field or edge devices.
2) Algorithms based on supervised learning components require extensive training with a large, comprehensive attack dataset which is sparsely available for CPS.
3) Detection strategies are methods to detect after the cyber-attack has impacted the physical system layer and is based on steady-state analysis of the system measurements.
4) Limited studies for CCA detection include coupling of both physical and cyber spaces to explicitly model the communication system vulnerabilities. Correlation of both spaces is crucial to establish the attack objective.

A. MOTIVATION

To address the aforementioned limitations, this paper aims to *i*) learn and correlate the relationship between different attack features to effectively detect the CCA goal, and *ii)* predict the attack events path to achieve a specific goal in order to prevent the CCA in the CPS environment.

In most studies, the dynamics of the power system are not modelled in the methodology. The motive of the adversary can be formulated as an attacker trying to find the unstable operating region in the grid to maximize the impact on the physical system. To understand the system behavior following cyber-attacks, the dynamics of a power system need to be incorporated to detect CCA and foresee the attacker's

path towards a specific goal. Dynamic security of the power system has been ensured traditionally through data-based approach, neural network-based learning, or defining stable operating region via nomograms [27], [28].

In this paper, a MBRL solution is proposed to detect CCA in the cyber layer. The scope of the work here is to develop an integrated CCA detection model for the distribution grid to learn the optimal policy of attack steps for identifying CCA objective where the learning algorithm is guided by the abductive reasoning criteria. The novelty of the proposed DCA is the MBRL designed as the integrated model of the environment and abductive reasoning based on power system dynamics to achieve policy optimization. The reasoning conditions utilize the operational behavior of the grid, providing an ''intelligent'' aspect to the detection strategy. The key contributions of this paper can be summarized as follows:

1) Intelligent IDS: The proposed detection system incorporates **abductive reasoning** criteria and dynamic simulation of the physical system to conclude CCA objectives and attack paths (Fig. 3C). Instead of solely relying on interacting with the environment for detection policy enhancement, both cyber layer alert information and physical layer data are leveraged to **correlate sequences of attack events**.
2) Distributed: The algorithm's distributed architecture ensures the **system-level detection** accuracy while **minimizing the communication overhead** by conducting localized detection at the node level.
3) Real-time detection: While RL typically lacks an offline mode, the proposed detection algorithm partitions it into offline learning (Fig. 3B and 3C) and online detection phases (Fig. 3D). This approach facilitates fast and resource-efficient online detection at the node level, beneficial for **nodes with limited computational capabilities**.
4) Comprehensive CPS: A **scalable CPS** incorporating the DNP3 communication protocol with the cyber system designed to quantify potential impacts of cyber-attacks and validate the proposed detection approach.

In the proposed algorithm, correlation between multiple attempted attack events in the cyber layer together with its intended physical impact serves as the basis for detecting the attacker's objective. This prevents any attack impact on the physical layer along with preventing further attack steps in the sequence by predicting them which makes it a proactive and preventive detection strategy.

The rest of the paper is organized as follows: Section II provides the CPS, attack model, and problem formulation. The design of the proposed MBRL solution based on abductive reasoning and the algorithm implementation are discussed in section III. The testbed setup developed at Virginia Tech (VT) is presented in section IV along with simulation results. Finally, section V provides the conclusion and future improvements.

II. BACKGROUND

In this section, the CPS model for distribution systems is presented, followed by introduction to the attack threat model, from the attacker's perspective and the problem formulation.

A. CYBER-POWER SYSTEM MODEL

The proposed CPS model has a decentralized structure, where each communication-capable node of the physical system, i.e., distribution feeder, has a corresponding node cyber layer, as shown in Fig. 1. The cyber layer is formed via the information and communications technology (ICT) path to which the physical layer component, e.g., the pole-top Feeder Remote Terminal Units (FRTU), send measurement packets. The ICT path is modeled as a *D/D/m/K*first-in-first-out (FIFO) queuing system, which is the fundamental communication mechanism, where the arrival time and service time of the data packets are deterministic (*D*). Every FRTU can hold K packets arriving at α packets/sec to be processed by $m$servers at a rate of β packets/sec [29]. The data from the node cyber layer is shared with the other nodes through a DNP3-based TCP/IP communication channel.

The measurement data packets are processed through Local analysis module (IDS). The detected individual attacks at the local level are sent to other nodes through DNP3-based communication [30]. DNP3 communication protocol is widely adopted for SCADA in the U.S. and beyond for monitoring and remote-control of field devices. At the node cyber layer, the measurement data packets after being screened by the Level 1 IDS generate attack alerts. The node measurements along with attack alerts are shared with other nodes through the communication module. The DNP3 Master-Outstation architecture is used, where the DNP3 data packets sent from outstations are received by the DNP3 master through periodic data polling, and then broadcasted to all other outstations.

B. ATTACK MODEL

Common cyber-attacks in the power grid such as FDI and replay attacks are considered here. It is assumed that attackers cannot directly access physical devices and have the capability to gain access remotely to the cyber layer of the communication-capable nodes. A series of attack events can be initiated either through compromised node(s) or by injecting false/modified data packets via man-in-the-middle attacks at the cyber layer. One or multiple adversaries can launch continuous attacks (sequential attacks at every time step) and irregular attacks (intermittent attacks at varying time steps) as shown in Fig. 2. Measurement data types that can be attacked in the CPS are given in Table 1.

Detection of the objective of a coordinated attack in the cyber layer enables taking mitigation steps which includes *i*) dropping attack data packets, *ii)* sending warning signals to the power components type in the predicted attack path, and *iii)* blocking of control command communication to the predicted attack nodes based on the CCA detection algorithm.

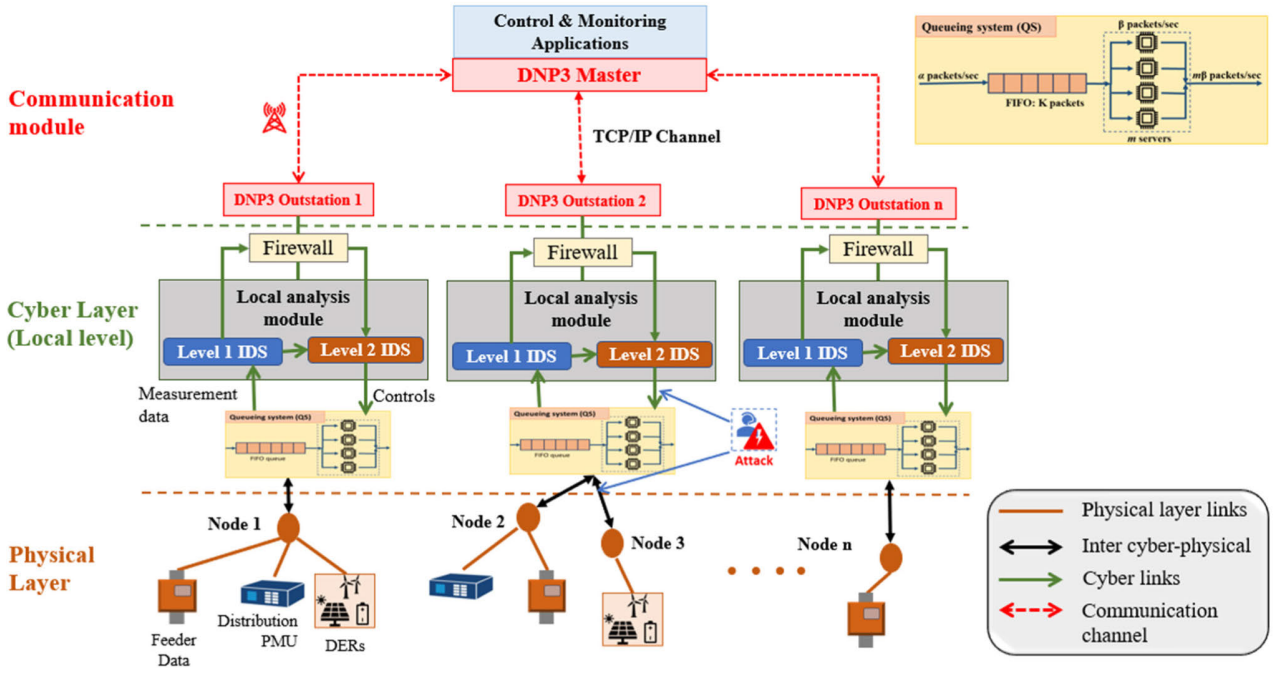

FIGURE 1. **Cyber-physical distribution system model with DNP3 communication.**

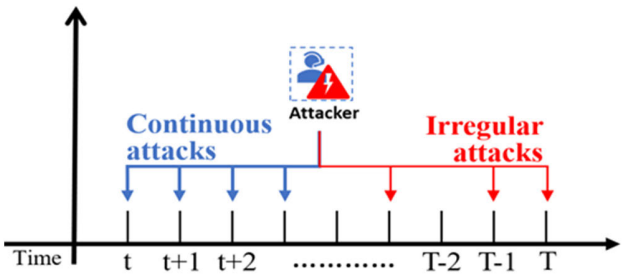

FIGURE 2. **Considered attack types.**

TABLE 1. **Possible attack data types.**

Attack component type	Attacked measurement data
Line	I, Switch
Load	P, Q, Switch
Generator	P, Q, Switch
Capacitor	Q, Switch
Breaker	Switch
Transformer	V, Switch

The attack model is designed from an attacker's perspective as the goal is to detect and predict the attack objective path. The intruder's motive can range from minor generator setpoint data tampering to cause load shedding to switching attacks to disrupt electricity service. The attacker's objective for a certain sequence of attack events s at time t, $A_o(s_t) \subset A_o$. Note: A_o is not an exhaustive list of attacker's intent list. Based on the distribution grid topology and available power components, A_o can have additional objectives for which corresponding reasoning conditions (discussed in Section III-C) should be defined for the learning algorithm to recognize the attack correlations. The CCA objective classification is defined as the following:

$$A_o = \begin{cases} 1, \text{complete power system outage} \\ 2, \text{loss of critical comp. \{gen, load, line\}} \\ 3, \text{remote switching command injection} \\ 4, \text{transformer tap tampering} \\ 5, \text{islanding} \\ 6, \text{instability due to data tampering} \\ 7, \text{line overloading} \\ 0, \text{inconclusive} \end{cases} \quad (1)$$

C. PROBLEM FORMULATION

To detect and prevent the attack impact on the physical system, the decentralized attack detection is formulated at the local analysis module in the cyber layer. The focus of the work is detecting CCA, which is designed as a systemwide Level 2 IDS in the cyber layer. However, it requires alerts to be generated by the Level 1 IDS upon detecting attempted attacks at the local node level.

Level 1 IDS detects attacks such as FDI and replay attacks at the node level. The cyber layer-based detection in [31] is adopted and modified to detect attacks and generate the attack alerts. This mechanism uses a combination of message authentication code (MAC) and timestamp to detect FDI and replay attacks, respectively. The measurement data along with its timestamp, at the physical node layer, is passed through a MAC algorithm to generate a unique MAC. The data, timestamp, and MAC are then sent as a data packet. The

intruder can attack data packets when it is sent over the cyber layer. If the data packet is modified, the calculated MAC at the cyber layer will not match the received MAC value and, therefore, an FDI attack alert is triggered. Similarly, a replay attack alert is created if the difference between the sent timestamp and the timestamp when data packet is received violates a tolerance value.

Attack alert generated at the nodes is expressed as $e \in \mathbb{R}^{11}$, and has the following features:

- Cyber attributes, CA ($\in \mathbb{R}^7$): source IP (attacker), destination IP (target node), attacker's username, attack power component ID, attack type, attack data (measurement datatype), timestamp
- Physical attributes, PA ($\in \mathbb{R}^4$): geographical location criticality, connection to critical loads, attack component type, connection status to the DERs

Detection based only on physical features such as the physical impact or changes in the measurements and breaker status provide relevant information for attack detection. However, this approach makes it a post-attack detection method and not a preventive method. Therefore, information from both cyber and physical domains are utilized in the proposed methodology, where the attempted attack sequence at the cyber layer is correlated with its simulated attack impact on physical domain along with other features to extract attack pattern. This information is further tested for the possibility of coordinated cyber-attack objective and predicting the attack trajectory.

At time t, there are 3 inputs to the Level 2 IDS of a node cyber layer for CCA detection: *i)* L_{loc}: attack alerts generated locally at the node from Level 1 IDS; *ii)* L_{ext}: external alerts that are generated locally at other nodes and communicated; and *iii)* L_{prev}: the sequence of alerts stored from the previous time window. Therefore, alerts available at time t are given by:

$$L^t_{avail} = L^t_{ext} U L^t_{prev} U L^t_{loc} \tag{2}$$

Note: $L^t_{prev} = L^{t-1}_{ext}$ U L^{t-1}_{prev} U L^{t-1}_{loc}. Based on L^t_{avail} over a sliding time window, the Level 2 IDS detects CCA at time t.

III. PROPOSED DCA ALGORITHM

A novel detection system designed for CCA is proposed to identify the correlation between cyber-attacks as shown in Fig. 3. The proposed strategy examines the discrete attack alerts generated from cyber layer to infer attacker's objective in the power grid. This is important in detecting attacks in the cyber layer before it affects the physical grid. The DCA algorithm i.e., Level 2 IDS is divided into the learning and online detection phases. After initial alert processing, an abductive reasoning based MBRL is performed in the learning phase. As the CCA detection in real-time is time-sensitive, a learned detection model (LDM) is utilized as opposed to complete real-time learning to reduce the computational overhead.

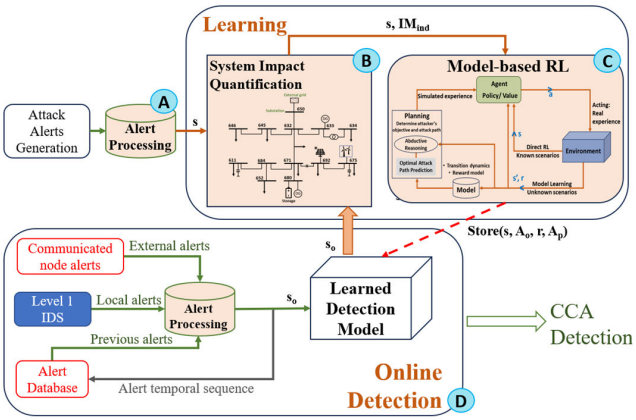

FIGURE 3. **Optimal CCA detection based on DCA algorithm.**

A. ALERT PROCESSING

Based on the attack attempts at different nodes in the cyber layer, a large volume of alerts may be generated at each time step. Hence, the extraction of the relevant attack information through alert processing includes alert clustering and temporal sequencing.

1) ALERT AGGREGATION AND CLUSTERING

Distributed alert aggregation ensures reduction in the computational and communication burden of detection algorithm through pre-processing alerts locally at the node level. It includes two actions:

i) Removing redundant alerts, i.e., same attack instance in L_{prev} U L_{loc} over the time window.
ii) Clustering alerts based on its cyber (CA) and physical (PA) features except time stamp.

The Density-Based Spatial Clustering and Application with Noise (DBSCAN) algorithm clusters the alerts and identify outliers, without pre-defining the number of clusters to be formed [32]. The IDS checks for similarities in the generated attack alerts dataset, where the number of attack clusters formed can vary between different time windows. The cluster is formed as an area with higher density, measured by examining if a point has MinPts (minimum number of points to form a cluster) within eps radius. Every data point P ∈ Dataset D, is tested to form an eps- neighborhood, given as:

$$eps_n(P) = \{Q \in D | dist(P, Q) \leq eps\} \tag{3}$$

To correlate the similarity between the points based on its non-binary features, the distance function is formulated as the cosine distance between two points P and Q, i.e.,

$$\begin{aligned} dist\,(P, Q) &= 1 - \text{cosine similarity}\,(P, Q) \\ &= 1 - \frac{|P \cap Q|}{\sqrt{|P|\,|Q|}} \\ &= 1 - \frac{\sum_{i=1}^{k} P_i \cdot Q_i}{\sqrt{\sum_{i=1}^{k} P_i^2} \cdot \sqrt{\sum_{i=1}^{k} Q_i^2}} \end{aligned} \tag{4}$$

The default MinPts value is three, i.e., at least three attack alerts are required to establish similarity. The dataset D is clustered into C_f clusters, where f is the number of clusters formed, based on the following condition:

$$\text{Clustering}, C_t = \begin{cases} C_f \text{ (core point)}, & \left|\text{eps}_n(P)\right| \geq \text{MinPts} \\ C_f \text{ (border point)}, & \left|\text{eps}_n(P)\right| < \text{MinPts} \\ \text{Noise}, & \text{otherwise} \end{cases} \tag{5}$$

In case of a sophisticated attack (multi-step CCA by multiple attackers), different attacker source IP addresses are in play. However, other features will be able to cluster them together as a preliminary step of the detection system. Hence, DBSCAN is used to cluster similar attack events in otherwise scattered and discrete alert log throughout different grid components.

2) ALERT TEMPORAL SEQUENCING

The multi-step attack can be coherent when the temporal relationship within the sequence of clustered attack events from different nodes is established. After clustering at time t, $C^t = \{C_1, C_2, C_3, \ldots\ldots, C_f\}$. If an earlier attack is a preliminary step for a later attack, the consequence of the earlier attack can reflect the existence of prerequisites for the later attack. Hence, the temporal sequence of alerts in each cluster is crucial to understand the attacker's motive. Nevertheless, a prerequisite is not considered as a mandatory condition. Any attack sequence S_f in cluster C_f, occurring in a temporal sequence of $[t_1 < t_2 < t_3 < \ldots.. < t_n]$, is defined as:

$$\forall C_f \in C^t, \quad S_f = [e_1 < e_2 < e_3 < \ldots.. < e_n] \tag{6}$$

where e_b $(1 \leq b \leq n)$ denotes the alert in cluster f, represented by a 11-tuple of cyber and physical attributes, $e_b^{C_f} = (CA^7, PA^4)$. The alert event sequence in S_f is arranged chronologically based on the timestamp cyber attribute, such that $e_i.\text{time} \leq e_j.\text{time}$, $(1 \leq i \leq j \leq n)$. Every cluster will have multiple attack sequences to be examined for a coordinated attack attempt, and the attack sequence set for cluster C_f is given as: $AS^t_{C_f} = \{S_1, S_2, \ldots, S_f\}$, where $S_m \subseteq S_f$, $(1 \leq m \leq f)$.

B. SYSTEM IMPACT QUANTIFICATION

Individual attacks are detected and alerts are generated at the cyber layer through Level 1 IDS, so the impact of attack attempt sequence is simulated. The system model is updated with topological change at time step t, given by $TP_t = \{sw_{comp}, load_{dem}\}$. The symbol $sw_{comp} \in \{0,1\}$ denotes the switching status of all the system components, where comp = {load, line, generator, transformer, capacitor, breaker}, and $load_{dem}$ is the real and reactive power demand of all the loads at time t. The power system model is defined and updated as:

$$M(s_t, TP_t) = IM_{ind} \tag{7}$$

System impact indicators IM_{ind}, where ind $\in$ (1,4), are based on the dynamic simulation of attempted attack sequence using the power system dynamic simulator, defined as:

i. Total loss of load, $IM_1 = \frac{P_{loss}}{P_{dem}} \in (0, 1)$
ii. Frequency deviation, $IM_2 = \frac{\Delta f}{f_{nom}}$, where $\Delta f = |f\text{-}f_{nom}|$
iii. Status of components (switching), $IM_3 = sw_{comp}$
iv. Voltage deviation, $IM_4 = \frac{|\Delta U|}{U_{nom}}$, where $|\Delta U| = |U_b\text{-}U_{nom}| \; \forall b \in \text{bus}$

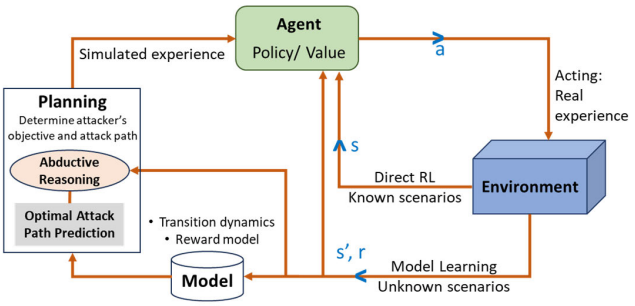

FIGURE 4. **Model-based RL component in DCA.**

C. MODEL-BASED REINFORCEMENT LEARNING

The proposed MBRL is formulated to identify the attacker's objective, and tentative attack path for CCA, shown in Fig. 4. The planning step provides a local solution over a subset of all the states available through interacting with the learned model. The RL learning agent stores a global solution in the form of policy function for the complete state space through interaction with the environment. The RL policy is improved until it reaches one of the attacker's objectives (final state) and learns the optimal attack path to reach objective A_o, given S_f.

The MBRL formulation is defined as a three tuple <S, A, r>, where S is the discrete state space (attack sequence), A is the discrete action space (possible attacks on available components) and r: S X A $\rightarrow$ R is the reward function. The objective of MBRL is to learn the optimal policy for CCA detection and is defined as a tuple ($\in \mathbb{R}^2$), i.e. <G> = {Attacker's objective A_o, Possible attack path A_p}. At t, the RL agent observes the state of the environment $s_t \in S$, which is the set of alert sequences in the clustered alerts at time t. Based on the RL policy π to maximize the long-term expected reward, the agent takes action $a_t \in A$, and receives reward r_t. RL agent aims to find an optimal policy $\pi*$ where the long-term expected reward is maximum, so the optimal value for state s, where discount factor, $\gamma \in (0, 1)$is:

$$V(s,\pi^*) = \max_a \{T(s, a, s') * (r(s, a) + \gamma V(s', \pi^*))\} \tag{8}$$

where T(s,a,s') is a transition probability function from state s to next state s'. The reward for the state-action pair is given as r (s, a) and V(s') is the expected cumulative reward.

MBRL learns a model that can simulate the environment dynamics, including the transition probabilities from one state to the next. In this case, the next state is deterministically obtained through dynamic simulation to learn multiple possible future trajectories to meet specific objectives. This

allows estimation of the expected cumulative rewards and optimization of the policy based on the observed deterministic dynamics.

The proposed DCA algorithm is modeled as an MBRL formulation which is concerned with finding an optimal policy that maximizes the cumulative reward to identify the attacker's objective based on attempted attack sequence and predict tentative optimal attack path to reach the terminal state.

1) REWARD FUNCTION

The RL agent maximizes the attacker's reward to capture the attack intent and attack path. The reward for an attacker is to cause maximum damage (P_{loss}) with less computational effort (number of attack steps) to avoid detection and impact the system. Hence, the reward for the agent is expressed as a reward function $r_t(s_t, a_t)$:

$$r_t = \frac{P_{loss}}{P_{dem}} + \frac{1}{n_{step}} + r_{obj} \tag{9}$$

where P_{loss} is the loss of load due to attack attempt, P_{dem} is the real power demand, and n_{step} is the number of alerts in the sequence. r_{obj} represents the reward obtained upon reaching the attack objective, which is a very high value upon satisfying the abductive reasoning condition and reaching the terminal state. During intermediate state transitions, $r_{obj} = 0$.

2) MODEL LEARNING

MBRL can perform efficiently when the model is learned sufficiently well with high-dimensional observables to plan and update the model with changes in power system dynamics. The proposed algorithm uses model learning where *i*) the interaction with the environment is simulated by the power system dynamic simulator, and *ii)* the policy is strengthened through abductive inference to identify A_o. The model learning includes learning the state transitions for specific objectives and the reward transitions to simulate environment interaction (eqn. (10)). Following the planning step, the experience is updated to optimize the policy function.

$$\text{Model}(s, a) = s', r \tag{10}$$

3) OPTIMAL ATTACK PATH PREDICTION THROUGH PLANNING

The planning step is formulated to predict possible next attack steps. Instead of random action trials to maximize the reward function, the planning steps are guided via abductive reasoning to predict the attack likelihood path A_p.

The decision making in MBRL is guided by the abductive reasoning criteria to reach a terminal state corresponding to the identified attack objective and maximized reward function. The agent then selects actions that are expected to lead to states and sequences of states that maximize the expected cumulative reward, improving the policy.

a: ABDUCTIVE REASONING FOR CAUSE-EFFECT CORRELATION

From the attacker's viewpoint, the attack sequence is a set of planning steps for a specific objective. Regardless of the attacker's level of system knowledge, the cause-effect relation between attack alert sequence (cause) and system impact (effect) establishes connection between different attack attempts. To classify the attacker's objective, a goal-oriented backward reasoning technique, abductive reasoning, is adopted to predict/ draw inference for a set of available observations, based on the domain-specific logic. In this case, considering the available attack alert sequence over the analyzed t_{window}, abductive reasoning provides a logically certain conclusion. The following conditions (one or multiple) need to hold true for specific attacker's objective (mapped from Eq. (1)) identification in an attack sequence:

- $(IM_1 = 1) \wedge (IM_3 \in comp_{load}) \Longrightarrow A_o(s_t) = 1$ (11)
- $\exists e_i \geq 3$ such that $(e_i.comp_{attack} \in comp_{critical}) \vee (IM_3 \in comp_{critical}) \Longrightarrow A_o(s_t) = 2$ (12)
- $\forall e_i$ such that $(e_i . comp_{attack} \notin comp_{critical}) \wedge (IM_3 \notin \text{comp critical}) \Longrightarrow A_o(s_t) = 3$ (13)
- $\forall e_i$ such that $(e_i . comp_{attack} \in transf_{tap}) \wedge (IM_4 \notin \text{limit}_{volt}) \Rightarrow A_o(s_t) = 4$ (14)
- $\forall e_i$ such that $e_i.comp_{attack} \in (PCC_{sw} \vee line(a,b))$ where $\forall$ line (a, b) such that $((\text{node a} \vee \text{node b}) \subseteq sw_{gen} = 1) \wedge ((IM_2(\text{node} \vee \text{node b}) \notin \text{limit}_{freq}) \vee (\text{system}_{unstable})) \Longrightarrow A_o(s_t) = 5$ (15)
- $\forall e_i$ such that $(e_i.comp_{attack} \in comp_{p\text{-}Q\ setpoint}) \wedge ((IM_2 \notin limit_{freq}) \vee (IM_4 \notin limit_{volt})) \Rightarrow A_o(s_t) = 6$ (16)
- $\forall e_i$ such that $(e_i.comp_{attack} \in comp\_dem) \wedge (IM_2 < \text{lower limit}_{freq}) \Longrightarrow A_o(s_t) = 7$ (17)

Based on the conditions given by Eq. (11)-(17), there can be three possibilities regarding attacker's objective:

i) Definite single attacker's objective (A_o): optimal attack path for the state s_t is reached and LDM is updated.
ii) Multiple objectives ($\subseteq A_o$) are satisfied: the selected objectives are examined through MBRL to predict optimal attack path to reach the most likely A_o.
iii) No single objective is fully satisfied: reasoning index utilized to find a subset of A_o to be examined.

In practice, the solution space of MBRL is reduced by use of abductive reasoning to conclude attacker's objective, greatly diminishing the computational burden. Further, the proposed reasoning index is used to narrow down the objective list A_o when there is a lack of alert observations in the available attempted attack sequence for a definitive objective conclusion. This further enables prediction of future attack sequence through faster learning.

1) Reasoning index:

Definite A_o is only feasible if the attacker has tried enough attack attempts by time t. However, incomplete alert sequence needs to be further examined. The incomplete alert sequence is characterized by an index to quantify the alert sequence approaching one or multiple A_o. This is an improvement against usual RL where all the possible actions are explored through value iteration before an optimal policy is obtained. The reasoning index of attack sequence S_f for objective condition x is expressed as $RI^x_{S_f} \in [0, 1]$ and is relative in nature:

$$RI^x_{S_f} = \frac{C_{evt}}{N_{evt}} = \begin{cases} 1, & \text{Attacker's obj identified} \\ > 0.5, & \text{Test } S_f \text{ for obj x} \\ < 0.5, & \text{Inconclusive : } A_o(s_t) = 0 \end{cases} \quad (18)$$

where, N_{evt}: total number of alert events in the sequence S_f, and C_{evt}: number of alert events (≥ 3) satisfying condition x. Note that at least 3 events in a sequence need to satisfy the reasoning condition to establish an attack goal.

MBRL is formulated to attain a definite A_o through multi-step lookahead of the future attack sequence to reach the attacker's likely objective. The DCA algorithm for learning the optimal policy to detect the CCA objective for a clustered attack sequence is given as Algorithm 1.

The RL agent starts with the attempted attack sequence, learns different attack path trajectories, and updates its policy to maximize the expected reward.

$$R_{\pi^*}(s_t) = \max_{\pi \in \Pi} R_\pi(s_t) \quad (19)$$

The RL agent learns over multiple episodes to predict the most probable attack path updated through value iteration until the iteration converges to the attack goal.

The proposed detection method is formulated to base its decision on the available alerts and the predicted optimal attack path. A failure to classify it to one of the categories can be a result of individual attacks on different components where the attack alerts are not correlated. The objective classification list is not exhaustive; new attack patterns that the algorithm learns through intermediate learning process can be used to expand the list of attacker's objectives.

D. ONLINE DETECTION MODEL

For real-time CCA detection, the online learned detection model is proposed to detect and predict CCA path as shown in Fig. 3 The alerts received, $L^t_{avail,}$ are processed in the Alert Processing component for generating multiple alert temporal sequences, $s_{o,}$ which is then checked for CCA. The LDM component checks the learned optimal policy for CCA detection. In case a new/ partial attack sequence is encountered, the state s_o is sent to the learning phase for learning the new state, updating the policy and the LDM.

E. ANALYSIS OF CONVERGENCE

The value function is a contraction mapping and by using the Banach fixed-point theorem, MBRL with value iteration will converge to the optimal policy. The following conditions ensure the convergence:

1. Finite state and action spaces
2. Deterministic transition dynamics of the system model are known and updated through dynamic simulation
3. Each alert sequence to be examined consists of at least 3 temporally sequenced attack events.

Contraction mapping is shown by the distance between $V_i(s)$ and $V_{i+1}(s)$, which is less than or equal to a constant k times the distance between $V_{i-1}(s)$ and $V_i(s)$ for any state s, where $k < 1$.

$$\begin{aligned} &|V_{i+1}(s) - V_i(s)| \\ &= \left| \max_a \left\{ \sum_{s'} T(s,a,s') * \left(R(s,a,s') + \gamma V_i(s') \right) \right\} \right. \\ &\quad \left. - \max_a \left\{ \sum_{s'} T(s,a,s') * \left(R(s,a,s') \gamma V_{i-1}(s') \right) \right\} \right| \\ &\leq \max_a \left\{ \sum_{s'} T(s,a,s') * \left| \gamma V_i(s') - \gamma V_{i-1}(s') \right| \right\} \end{aligned} \quad (20)$$

Algorithm 1 MBRL (s, IM_{ind})

```
Initialization
    Initialize V(s,a), M(s,a), and learning parameters
Select possible A_o
    Check Abductive Reasoning conditions for A_o
    if len (A_o) = 1 → Definite A_o then
        Compute r and Update LDM (s, A_o, r)
        break
    if length (A_o) = 0 → No single A_o then
        calculate RI^x for s and list of possible A_o
    else→ Multiple A_o
        list of A_o satisfied
    end if
    Define the possible action set A for attacker objective A_o
for each episode do
    for each Learning step do
        State S ← available state s
        Attack path P ← [S]
        Acting: Take action a ∈ A, receive reward r and next state s'
        Update value function using eqn. (8):
        V(s,a) ← max_a ← {r(s, a) + γV(s', π*))
        Model Learning: Update Model (s,a) ← s', r
        for N Planning step do
            s, a ← previously observed (s, a)
            s', r ← Model (s,a)
            Update V(s,a) ← max_a ← {r(s, a) + γV(s', π*))
        end for
        S ← next state s'
        Update possible actions A ← A - a
        if selected A_o is reached then break
        end if
        Attack path P ← Append S
    end for
    Optimal path list ← Attack path P
Select the attack path P from Optimal path list with highest r using eqn. (19)
end for
```

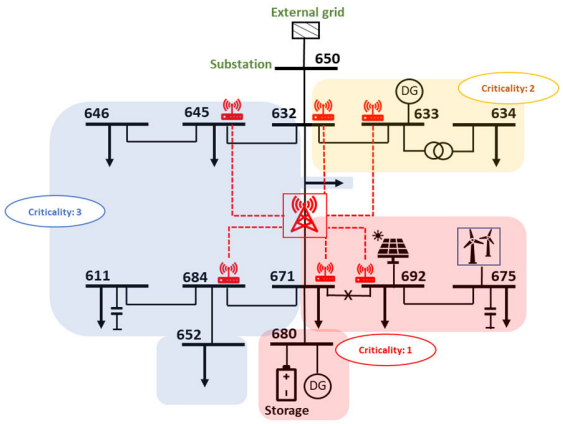

FIGURE 5. **IEEE 13-node feeder with added DERs.**

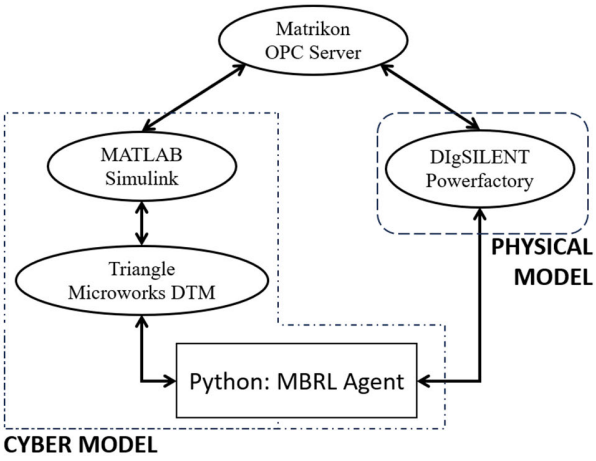

FIGURE 6. **Cyber-power system simulation setup for DCA.**

Using the sum of transition probabilities for state s and action a is equal to one:

$$|V_{i+1}(s) - V_i(s)| \leq \gamma * k * |V_i(s) - V_{i-1}(s)|$$

One can find a constant $k < 1$ by taking the maximum of the sum of transition probabilities times the distance between the value functions over all states and actions.

IV. SIMULATION AND RESULTS

A. VT CPS TESTBED

There is limited availability of cyber-physical data in a distribution network pertaining to cybersecurity concerns. This includes measurements from feeder nodes (such as voltages, currents, switch status, and power) and alert information from the cyber system. Hence, a realistic testbed with physical-cyber-communication module is developed at VT to simulate and study attack impacts and test DCA.

The IEEE 13-node test feeder is modified to include DERs and is used as a microgrid test system, shown in Fig. 5. All components in the system have remote-control capabilities and the microgrid is operating in an islanded mode. Based on the placement of critical components (Loads: 670, 671, 675; Lines: 632-645, 671-680, 671-684; Generators: DG_633, DG_680, PV), the network is divided into different areas of criticality (physical attribute) with Criticality:1 being the most critical.

The attack training dataset is generated through mimicking the attacker's behavior by initiating multiple attacks (both continuous and irregular attacks) at the cyber layer. These attacks on the system components are done both randomly and in a strategically curated fashion. The Level 1 IDS at the cyber layer of developed CPS generates the timestamped attack alerts. For the offline Learning phase to learn the optimal policy, the generated attack alerts of all the nodes over a time window (L^t_{avail}) is assessed by the proposed learning-based detection method to correlate multiple attempted attacks and learn attack paths to achieve specific attack objective. The feeder is the environment for MBRL component in DCA. The time step of the power system dynamic simulation, $\Delta t = 0.01$ s. The simulation time window (t_{win}) is 1s and the sliding window size is 0.1s. For real-time detection, the developed co-simulation platform includes a power distribution system with static and dynamic system models, implemented in DIgSILENT Powerfactory, and the cyber layer in MATLAB Simulink. Real-time data exchange is through Matrikon OPC server with DNP3 communication modeled in Triangle Microworks DTM software and Python programming is used to model RL agent to detect CCA. The interaction among these software components for CPS is shown in Fig. 6. The snapshot of the test simulation setup in DTM software for online detection of CCA is shown in Fig. 7. DNP Master periodically polls measurement data from all node Outstations during normal operation. At the cyber layer, data packets are examined, and individual attack alerts are generated and displayed in "Attack Alerts !." For a t_{win}, the DCA algorithm detects attacker objective and predicts possible attack path as shown in "CCA Detection Report." The attack packets are dropped, and warning signals are sent to the predicted attack components type. In Fig. 7., based on the first three attack alerts, the Optimal Attack Path is predicted to be a DER, warning signals (shown as blue box) are sent to all the DERs present in the system, prior to the occurrence of other attack alerts shown.

To validate the proposed algorithm, two CCA scenarios (Table 2) are tested on the CPS testbed to show the variation in attacker's attack objective and efficient detection results.

1) SCENARIO 1: MULTI-STAGE CONTINUOUS ATTACK

A total of 56 individual alerts are generated in the first t_{win}, T_{win-1} (t_0-t_1). 7 alerts are clustered in Alert processing: [t_{a1}, t_{a2}, t_{a5}, t_{a6}, t_{a9}, t_{a10}, t_{a12}], shown in Fig. 8, as they are attempted by the same attacker (cyber attribute: Source IP). Based on the available clustered alert sequence at T_{win-1} and the system impact quantification, MBRL limits its action set defined by abductive reasoning to check cause-effect

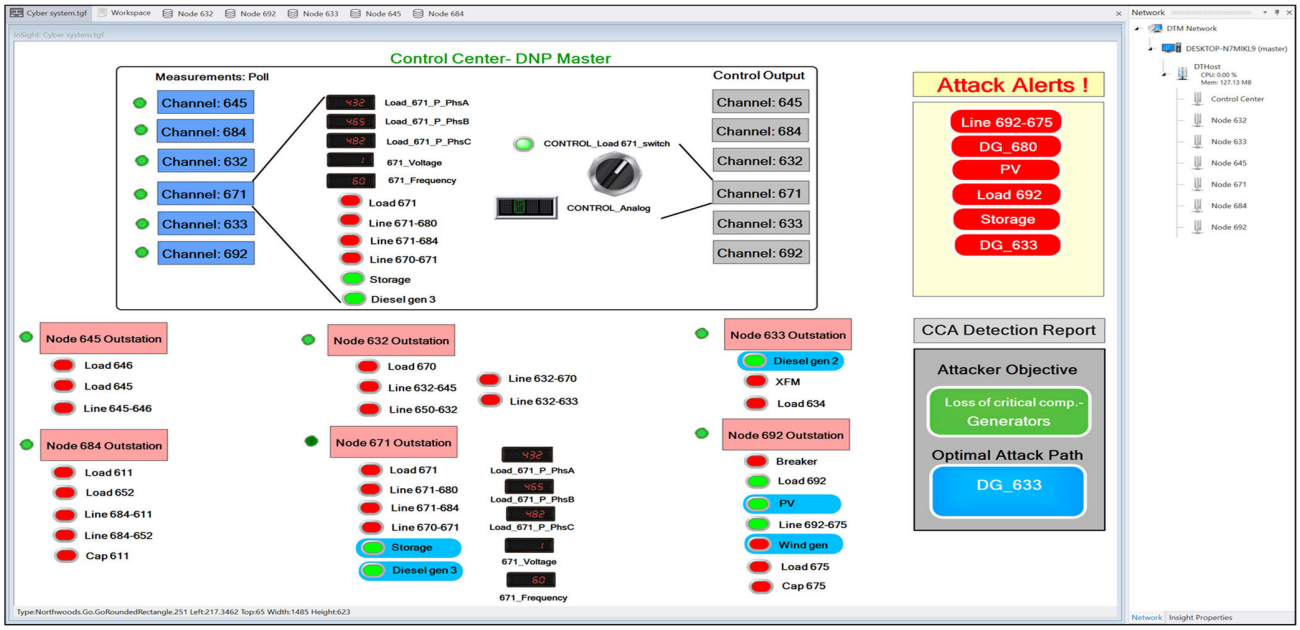

FIGURE 7. **Virginia Tech Cybersecurity Testbed for IEEE 13-node test feeder.**

TABLE 2. **Attempted attack event sequence of test scenarios.**

SCENARIO 1			SCENARIO 2		
Time	Target	Attack data	Time	Target	Attack data
t_{a1} (5:03:04)	Line 671-684	I	t_{a2} (6:01:02)	Line 692-675	Switch
t_{a2} (5:03:07)	PV	Switch	t_{a4} (6:01:29)	Load 692	P
t_{a5} (5:03:09)	Load 692	Switch	t_{a9} (6:01:33)	Cap 675	Q
t_{a6} (5:03:54)	DG 680	Switch	t_{a13} (6:01:68)	Line 632-633	Switch
t_{a9} (5:03:59)	Load 670	P	t_{a14} (6:01:97)	Breaker 671-692	Switch
t_{a10} (5:03:73)	Load 611	Q	t_{a18} (6:02:23)	Load 670	Q
t_{a12} (5:03:98)	Line 684-652	Switch	t_{a22} (6:02:54)	Line 645-646	I
t_{a13} (5:04:03)	Load 692	Switch	t_{a27} (6:02:59)	Line 671-680	Switch
t_{a14} (5:04:06)	Line 632-645	Switch			
t_{a17} (5:04:07)	Load 675	P			
t_{a18} (5:04:10)	Line 632-633	Switch			

correlation for specific objectives $\subset A_o$. The optimal RL policy detects that the attacker's objective is to disconnect critical components ($A_o = 2$) and with disconnection of one single critical component, i.e., A_p: DG_633 (less number of steps from attacker), the attacker's reward is the maximum. However, as there is more attack alert information available in the next sliding time window $t_{0.1}$-$t_{1.1}$, i.e., T_{win-2}, the new clustered attack sequence: [t_{a6}, t_{a9}, t_{a10}, t_{a12}, t_{a13}, t_{a14}, t_{a17}, t_{a18}] is examined. From the learned optimal policy, the attacker is trying to create a complete power system outage

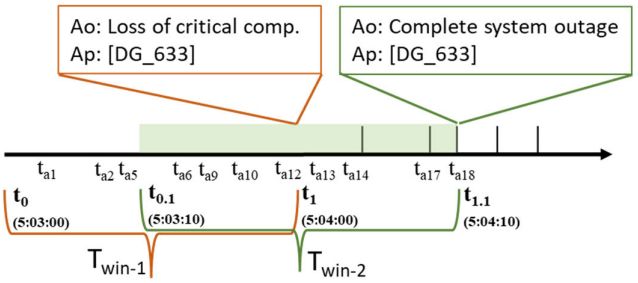

FIGURE 8. **Detection result of Scenario 1.**

($A_o = 1$). The objective identification is based on available attack alerts, so as there are more evidence over sliding time windows, the DCA learns the optimal policy to conclude attacker's objective. The system impact quantification through dynamic simulation of the attempted attack sequence in the cyber layer is shown in Figs. 9(a) and 9(b). As CCA is detected at the cyber layer preventing physical system impact, Figs. 9(c) and 9(d) show that the system frequency and node voltages are maintained within acceptable operational limits.

2) SCENARIO 2: MULTI-STAGE IRREGULAR ATTACK

This scenario shows the DCA efficiency and attack path prediction for discontinuous attacks over multiple time windows. In T_{win-1}, the attack alerts: [t_{a2}, t_{a4}, t_{a9}, t_{a13}, t_{a14}] are clustered together out of 48 individual alerts based on the geographical location criticality (physical attribute). Based on available alerts and its probable impact quantification through dynamic simulations, through reasoning-based planning steps, MBRL learns the attack path. The learned optimal

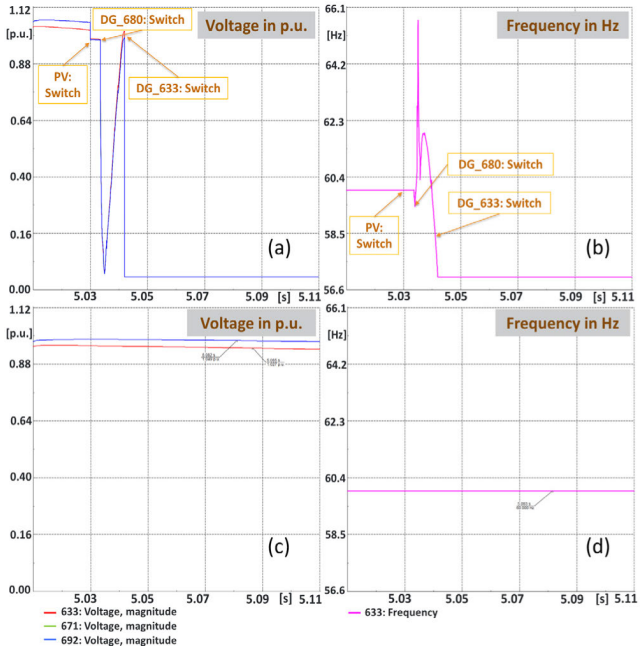

FIGURE 9. Impact on power system: (a) - (b) DCA algorithm's Simulated Scenario 1 with predicted Ap to visualize impact on the voltage and system frequency in Level 2 IDS, and (c) - (d) Actual impact on the power grid after Scenario 1 is detected in the cyber layer.

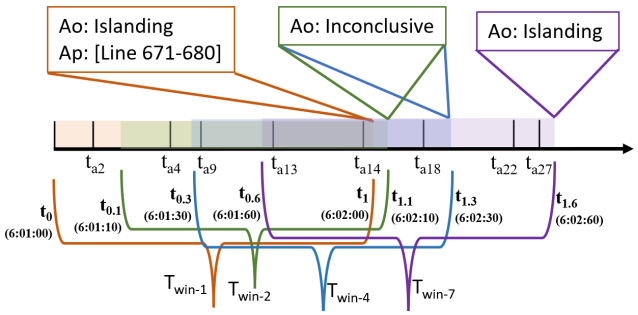

FIGURE 10. Detection result of Scenario 2.

policy detects the attacker's objective to be islanding ($A_o = 5$), shown in Fig. 10. Optimal path A_p prediction for maximum reward, i.e., multiple islands with maximum loss of load are by remote switching one component, i.e., Line 671-680. This will result in two islands: Island 1 (Node 633 and Node 634 where Load 634 demand is met by DG_633) and Island 2 (Node 680 with DERs present). As the sliding time window clusters additional attempted attacks, the sequence of events does not have a definitive conclusion ($A_o = 0$) over T_{win-2} - T_{win-6}. The alert database is a memory for multiple time windows. In T_{win-7}, with clustered attack alerts: [t_{a13}, t_{a14}, t_{a18}, t_{a22}, t_{a27}], the attempted attack alert is t_{a27} which was the predicted attack path from T_{win-1}. Hence, the DCA algorithm learns the policy over multiple time windows and correlates the attacker's objective.

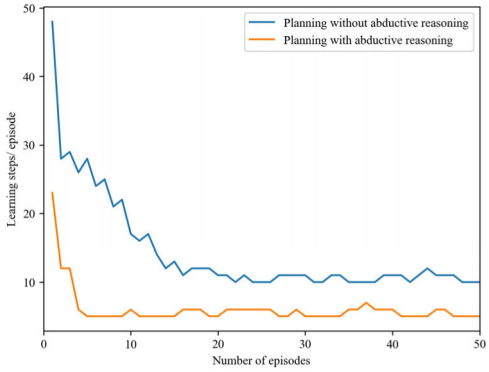

FIGURE 11. Average learning curve of MBRL agent with 20 planning steps (N).

TABLE 3. CCA detection performance.

	Same initial conditions	New system configuration
Learning + LDM (s)	0.089	>> 0.1
Only LDM (s)	0.042	0.056

B. LEARNING CURVES IN DCA

The DCA algorithm aims to trace the most probable attack path A_p from the attacker's perspective, given an attack sequence. The planning component in the MBRL is guided by abductive reasoning to reach the terminal state (objective A_o) in fewer learning steps for faster learning. For a fixed number of planning step N, Fig. 11 shows the number of learning steps in each episode taken by the RL agent. It is averaged over 20 runs of the same attack data and is tested with and without abductive reasoning. With abductive reasoning, the planning agent took about 5 episodes to attain optimal performance with a smaller number of learning steps. The agent took 20 episodes to reach the optimal policy without reasoning as it had to go through a much larger action set.

C. ONLINE DETECTION PERFORMANCE

The CCA online detection involves clustering the alerts and checking each available alert sequence against the LDM for learned attack paths. In absence of an already learnt attack path, the attack sequence goes through the learning phase and the detection performance is impacted. Table 3 summarizes the online detection performance of the DCA algorithm. Different attack sequences are tested over multiple time windows to attain the average detection time. In case of the same system configuration over sliding time windows, if the attack sequence to be tested is already learnt, the online detection model goes through only LDM which takes ~0.042 sec to detect. In case of a new/ partial learnt attack sequence, the attempted attack sequences are sent to the MBRL-based learning where the agent will trace the amount of known information through its value function update and learn the attack sequence impact from that state. It then goes through the planning phase of MBRL to predict an optimal attack path

and update its policy. The new/ partial learnt attack sequence takes ∼0.089 sec to detect CCA. This implies that before the next time window, the algorithm can detect the CCA efficiently provided the system configuration is unchanged. However, with an altered system configuration, it takes well above the moving window size to learn the optimal policy for detecting CCA. If the changed system configuration is part of LDM, the attack sequence can be tested to detect an attack objective in ∼0.056 s. Therefore, extensive learning with different system configurations in offline mode provides LDM with policy that can be improved further during online detection.

V. CONCLUSION

The proposed IDS depends on both the cyber layer alert information and physical layer data to correlate the sequence of attack events in the cyber layer. It extracts the physical component target and analyzes the probable system impact, so irrespective of the type of distinct initial attack, the algorithm is adequate to detect the attack objective and predict the attack path in real-time. This can provide operators with evidence to take action to stop future attacks. This proposed strategy focuses on interpretable decision-making as the cause-effect correlation is considered in MBRL to detect CCA, instead of learning the model as a black box. MBRL enhances interpretability by explicitly modeling environment dynamics, enabling simulation and planning, providing traceability of decisions, and transparency in learning updates. These aspects aid in understanding and justifying decision-making processes in identifying the CCA objective. Scalability is ensured through updating the considered power system model to update the changes in system states. However, with very large networks, the time for learning process increases.

In online detection, when the LDM encounters a new attack sequence, it goes through the Learning phase, thus increasing the detection time which can result in substantial delays in larger networks. Future work is needed to address this issue, along with algorithm enhancement to consider contaminated state measurements. The online detection phase utilized a learned detection model which works like a look-up table that has the state action mapping for different state transitions. The quantification of the memory overhead for the DCA algorithm requires hardware-in-the-loop testing. DCA is dependent on the availability of grid information such as grid topological data, measurement data, and cyber layer information. Dealing with the limited availability of such real-time information in the distribution system is a hurdle to be overcome.

REFERENCES

[1] M. Baezner and P. Robin, "Stuxnet," ETH Zürich, Zürich, Switzerland, Oct. 2017, pp. 1–14.

[2] T. Plėta, M. Tvaronavičienė, S. D. Casa, and K. Agafonov, "Cyber-attacks to critical energy infrastructure and management issues: Overview of selected cases," *Insights Regional Develop.*, vol. 2, no. 3, pp. 703–715, Sep. 2020.

[3] Z. Dehlawi and N. Abokhodair, "Saudi Arabia's response to cyber conflict: A case study of the shamoon malware incident," in *Proc. IEEE Int. Conf. Intell. Secur. Informat.*, Jun. 2013, pp. 73–75.

[4] J. W. Goodell and S. Corbet, "Commodity market exposure to energy-firm distress: Evidence from the colonial pipeline ransomware attack," *Finance Res. Lett.*, vol. 51, Jan. 2023, Art. no. 103329.

[5] D. U. Case, "Analysis of the cyber attack on the Ukrainian power grid," *Electr. Inf. Sharing Anal. Center*, vol. 388, nos. 1–29, p. 3, 2016.

[6] J. Valenzuela, J. Wang, and N. Bissinger, "Real-time intrusion detection in power system operations," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1052–1062, May 2013.

[7] C.-C. Liu et al., "Cyber-physical system security of stribution systems," *Found. Trends® Electr. Energy stems*, vol. 4, no. 4, pp. 346–410, 2021.

[8] J. Hong, C.-C. Liu, and M. Govindarasu, "Integrated anomaly detection for cyber security of the substations," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1643–1653, Jul. 2014.

[9] Y. Zhang, L. Wang, W. Sun, R. C. Green II, and M. Alam, "Distributed intrusion detection system in a multi-layer network architecture of smart grids," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 796–808, Dec. 2011.

[10] M. Parvania, G. Koutsandria, V. Muthukumary, S. Peisert, C. McParland, and A. Scaglione, "Hybrid control network intrusion detection systems for automated power distribution systems," in *Proc. 44th Annu. IEEE/IFIP Int. Conf. Dependable Syst. Netw.*, Jun. 2014, pp. 774–779.

[11] D. An, Q. Yang, W. Liu, and Y. Zhang, "Defending against data integrity attacks in smart grid: A deep reinforcement learning-based approach," *IEEE Access*, vol. 7, pp. 110835–110845, 2019.

[12] S. Paul and Z. Ni, "Vulnerability analysis for simultaneous attack in smart grid security," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Apr. 2017, pp. 1–5.

[13] Z. Ni, S. Paul, X. Zhong, and Q. Wei, "A reinforcement learning approach for sequential decision-making process of attacks in smart grid," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2017, pp. 1–8.

[14] M. Liu, C. Zhao, Z. Zhang, R. Deng, P. Cheng, and J. Chen, "Converter-based moving target defense against deception attacks in DC microgrids," *IEEE Trans. Smart Grid*, vol. 13, no. 5, pp. 3984–3996, Sep. 2022.

[15] C.-C. Sun, J. Hong, and C.-C. Liu, "A coordinated cyber attack detection system (CCADS) for multiple substations," in *Proc. Power Syst. Comput. Conf. (PSCC)*, Jun. 2016, pp. 1–7.

[16] L. Wang et al., "Method for extracting patterns of coordinated network attacks on electric power CPS based on temporal-topological correlation," *IEEE Access*, vol. 8, pp. 57260–57272, 2020.

[17] J. Appiah-Kubi and C.-C. Liu, "Decentralized intrusion prevention (DIP) against co-ordinated cyberattacks on distribution automation systems," *IEEE Open Access J. Power Energy*, vol. 7, pp. 389–402, 2020.

[18] N. Sahani, R. Zhu, J.-H. Cho, and C.-C. Liu, "Machine learning-based intrusion detection for smart grid computing: A survey," *ACM Trans. Cyber-Phys. Syst.*, vol. 7, no. 2, pp. 1–31, Apr. 2023.

[19] L. Wang et al., "Coordinated cyber-attack detection model of cyber-physical power system based on the operating state data link," *Frontiers Energy Res.*, vol. 9, Apr. 2021, Art. no. 666130.

[20] N. Bhusal, M. Gautam, R. M. Shukla, M. Benidris, and S. Sengupta, "Coordinated data falsification attack detection in the domain of distributed generation using deep learning," *Int. J. Electr. Power Energy Syst.*, vol. 134, Jan. 2022, Art. no. 107345.

[21] J. Appiah-Kubi and C.-C. Liu, "Cyberattack correlation and mitigation for distribution systems via machine learning," *IEEE Open Access J. Power Energy*, vol. 10, pp. 128–140, 2023.

[22] Z. Ni and S. Paul, "A multistage game in smart grid security: A reinforcement learning solution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2684–2695, Sep. 2019.

[23] Z. Wang, H. He, Z. Wan, and Y. Sun, "Coordinated topology attacks in smart grid using deep reinforcement learning," *IEEE Trans. Ind. Informat.*, vol. 17, no. 2, pp. 1407–1415, Feb. 2021.

[24] M. Liu et al., "Enhancing cyber-resiliency of DER-based smart grid: A survey," *IEEE Trans. Smart Grid*, vol. 15, no. 5, pp. 4998–5030, Sep. 2024.

[25] M. Tarle, M. Björkman, M. Larsson, L. Nordström, and G. Ingeström, "A world model based reinforcement learning architecture for autonomous power system control," in *Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm)*, Oct. 2021, pp. 364–370.

[26] S. Totaro, I. Boukas, A. Jonsson, and B. Cornélusse, "Lifelong control of off-grid microgrid with model-based reinforcement learning," *Energy*, vol. 232, Oct. 2021, Art. no. 121035.

[27] Y. Mansour, A. Y. Chang, J. Tamby, E. Vaahedi, B. R. Corns, and M. A. El-Sharkawi, "Large scale dynamic security screening and ranking using neural networks," *IEEE Trans. Power Syst.*, vol. 12, no. 2, pp. 954–960, May 1997.
[28] H. Sarmiento, G. Pampin, R. Barajas, R. Castellanos, G. Villa, and M. Mirabal, "Nomograms for assistance in voltage security visualization," in *Proc. IEEE/PES Power Syst. Conf. Expo.*, Mar. 2009, pp. 1–6.
[29] A. Stefanov, C.-C. Liu, M. Govindarasu, and S.-S. Wu, "SCADA modeling for performance and vulnerability assessment of integrated cyber-physical systems," *Int. Trans. Electr. Energy Syst.*, vol. 25, no. 3, pp. 498–519, Mar. 2015.
[30] P. Wlazlo et al., "Man-in-the-middle attacks and defence in a power system cyber-physical testbed," *IET Cyber-Phys. Syst., Theory Appl.*, vol. 6, no. 3, pp. 164–177, Sep. 2021.
[31] A. K. Jain, N. Sahani, and C.-C. Liu, "Detection of falsified commands on a DER management system," *IEEE Trans. Smart Grid*, vol. 13, no. 2, pp. 1322–1334, Mar. 2022.
[32] I. Cordova and T.-S. Moh, "DBSCAN on resilient distributed datasets," in *Proc. Int. Conf. High Perform. Comput. Simulation (HPCS)*, Jul. 2015, pp. 531–540.

NITASHA SAHANI (Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from Virginia Tech, Blacksburg, USA, in 2019 and 2024, respectively.

Her research interests include communication and cybersecurity of smart grids, microgrid co-simulation, electric vehicle charging schedules, and applied optimization for power systems applications.

CHEN-CHING LIU (Life Fellow, IEEE) was a Faculty Member of the University of Washington, Iowa State University, University College Dublin, Ireland, and Washington State University, from 1983 to 2017. He is currently an American Electric Power Research Professor and the Director Emeritus of the Power and Energy Center, Virginia Tech. He is the U.S. Member of the CIGRE Study Committee D2. He is a member of the U.S. National Academy of Engineering. He received the IEEE Third Millennium Medal in 2000 and the Power and Energy Society Outstanding Power Engineering Educator Award in 2004. He received the Attwood Associate and Dale Douglass Award from the U.S. National Committee of CIGRE.