Received 17 June 2022; revised 13 October 2022; accepted 19 December 2022.
Date of publication 12 January 2023; date of current version 24 January 2023.

Digital Object Identifier 10.1109/OAJPE.2023.3236429

Cyberattack Correlation and Mitigation for Distribution Systems via Machine Learning

JENNIFER APPIAH-KUBI (Student Member, IEEE)
AND CHEN-CHING LIU (Life Fellow, IEEE)
The Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute
and State University, Blacksburg, VA 24061 USA
CORRESPONDING AUTHOR: J. APPIAH-KUBI (jennifera@vt.edu)

This work was supported in part by the U.S. National Science Foundation under Grant 1837359; and in part by the projects under Grant PR6ZBSTA and Grant 5GPG sponsored by Commonwealth Cyber Initiative (CCI), State of Virginia.

ABSTRACT Cyber-physical system security for electric distribution systems is critical. In direct switching attacks, often coordinated, attackers seek to toggle remote-controlled switches in the distribution network. Due to the typically radial operation, certain configurations may lead to outages and/or voltage violations. Existing optimization methods that model the interactions between the attacker and the power system operator (defender) assume knowledge of the attacker's parameters. This reduces their usability. Furthermore, the trend with coordinated cyberattack detection has been the use of centralized mechanisms, correlating data from dispersed security systems. This can be prone to single point failures. In this paper, novel mathematical models are presented for the attacker and the defender. The models do not assume any knowledge of the attacker's parameters by the defender. Instead, a machine learning (ML) technique implemented by a multi-agent system correlates detected attacks in a decentralized manner, predicting the targets of the attacker. Furthermore, agents learn optimal mitigation of the communication level through Q-learning. The learned attacker motive is also used by the defender to determine a new configuration of the distribution network. Simulations of the technique have been performed using the IEEE 123-Node Test Feeder. The simulation results validate the capability and performance of the algorithm.

INDEX TERMS Intrusion detection, cyber security, anomaly detection, q-learning, reinforcement learning, multi-agent systems, entropy, distribution automation, distribution reconfiguration.

NOMENCLATURE

$\mathcal{S}$	Set of source nodes.
$\mathcal{D}$	Set of demand nodes.
$\mathcal{N}$	Set of nodes in distribution network.
$\mathcal{E}$	Set of distribution lines.
$\mathcal{D}_s$	Set of demand nodes whose load are connected through remote-control switches.
$\mathcal{E}_s$	Set of lines with remote-control disconnect switches.
$\mathcal{D}_{sp}$	Set of demand node remote-control switches whose agents have shut down remote control capability.
$\mathcal{E}_{sp}$	Set of line remote-control switches whose agents have shut down remote control capability.
$\overline{V}$	Upper voltage limit.
$\underline{V}$	Lower voltage limit.
U_i^ϕ	Square of voltage magnitude at phase ϕ of node i.
P_i^ϕ	Active demand at phase ϕ of demand node i following attack.
P_{ij}^ϕ	Active power flow on phase ϕ of line (i, j) following attack.
Q_i^ϕ	Reactive demand at phase ϕ of demand node i following attack.
Q_{ij}^ϕ	Reactive power flow on phase ϕ of line (i, j) following attack.
P_{ic}^ϕ	Current active demand at phase ϕ of node i prior to attack.
Q_{ic}^ϕ	Current reactive demand at phase ϕ of node i prior to attack.
P_{gi}^ϕ	Active power injection on phase ϕ of node i.
Q_{gi}^ϕ	Reactive power injection on phase ϕ of node i.
θ_i^ϕ	Angle of load at phase ϕ of node i.

q_i	Attack quality of demand node i.
κ_i	Criticality of load at demand node i.
Φ_i	Set of phases at node i.
Φ_{ij}	Set of line phases between node i and node j.
r^{ij}	Resistance matrix for lines between node i and node j.
x^{ij}	Reactance matrix for lines between node i and node j.
R	Attacker's monetary reward per kW power disrupted.
B	Attacker's monetary budget.
C_F	Fixed monetary cost of attack.
$C_{e_t}^{ij}$	Attacker's monetary variable cost to toggle remote-control switch of line (i, j).
$C_{e_k}^{ij}$	Attacker's monetary variable cost to perform denial of service attack on remote-control switch of line (i, j).
C_d^i	Attacker's monetary variable cost to toggle remote-control switch connecting load at demand node i.
e_{ij}^t	Binary variable indicating that remote-control switch on line (i, j) is to be toggled.
e_{ij}^k	Binary variable indicating that remote-control switch on line (i, j) is to be kept in the current state.
z_i	Binary state of remote-control switch that connects load at demand node i following attack.
s_{ij}	Binary state of remote-control switch on line (i, j) following attack.
s_{ij}^c	Current binary state of remote-control switch on line (i, j).
y_{ij}	Binary operator-controlled state of remote-control switch on line (i, j) not selected for attack.
w_{ij}	Binary variable indicating power flow direction on line (i, j).
F_{ij}	Power flow limit of line (i, j).
c_t	Default NIDS threshold for detecting attack c.
c_n	New NIDS threshold for detecting attack c following alert receipt.
δ	Scaling parameter used in operator's reconfiguration.
W	Sliding window of time within which anomaly is monitored.
T	Time period within which communication level mitigation is enforced.
ρ	Attack likelihood index.
$\mathcal{N}_L$	Set of unique communication network types in the set of received alerts.
$\mathcal{F}_L$	Set of unique firmware types in the set of received alerts.
$\mathcal{N}_N$	Set of unique communication network types of all agents in the distribution system.
$\mathcal{F}_N$	Set of unique firmware types of all agents in the distribution system.
u^n	Communication network type of an agent.
u^f	Firmware type of an agent.
ψ	Similarity index for operational properties.
K	Number of alerts to activate physical mitigation.
s_t	Reinforcement learning state at training time step t.
a_t	Action performed during training time step t.
r_t	Reward given to reinforcement learning agent at training time step t.
α	Learning rate of reinforcement learning agent.
ϵ	Exploration rate of reinforcement learning agent.

I. INTRODUCTION

WITH the integration of advanced communication technology, the power grid is increasingly remotely monitored and controlled. Nevertheless, the advancement has also made the smart grid more vulnerable to cyberattacks. In December 2015, six distribution utilities in Ukraine suffered cyberattacks. The ensuing outage affected about 225,000 customers [1].

Significant research has been conducted in the area of distribution system cybersecurity, and several techniques have been proposed for different applications.

A. RELATED WORK

As a cyber-physical system, the power grid, including distribution systems, is vulnerable to various forms of cyberattacks [2], [3] such as false data injection attacks [4] and load altering attacks [5]. These attacks are threats to the stability and control of the target power grid. However, since there are cyber intrusion detection techniques in place, say, those associated with state estimation and bad data detection, such attacks are covertly and stealthily launched, making them difficult to execute. Another attack type that may well effect dire consequences on the power grid is the control signal attack, including the direct switching attack. In control signal attacks, the attacker aims to gain direct control over the physical device, and the attacks are often not covert [2].

By direct switching attacks, switches and circuit breakers connecting power system equipment such as lines, load, and generators are toggled. The attacks tend to be coordinated as multiple elements in the grid need to be attacked to achieve the objective of the attacker on the radial distribution network. In [6], a set of decentralized algorithms are presented to detect man-in-the-middle attacks on a distribution system. The algorithms aim to prevent direct switching of circuit breakers and tampering with relay settings that could lead to voltage violations and inconsistent protections settings. In [7], a decentralized algorithm is put forward to address coordinated switching attacks on the power distribution system. The algorithm predicts the targets of a coordinated cyberattack ahead of the attacker, and determines mitigation strategies. The concept of attack target prediction is explored in [8]. Here, attack templates are used to pre-compute substation correlation sets for attacks. When an attack is detected at a substation, the closest fitting set is selected and protected. It is noted that the technique in [8] uses a centralized architecture.

The direct switching attack is also studied using interdiction models of the interactions between the attacker and system operator in a bi- or tri-level optimization problem [3]. The tri-level optimization problem is known as the defender-attacker-defender (DAD) model whereas the bi-level model is an attacker-defender (AD) model. In [9] a DAD model is presented that includes the defender's planning stage hardening decisions, the attacker's coordinated decisions to maximize damage, and the defender's attack response, such as distribution network reconfiguration (DNR) and optimal DG islanding. The DAD model in [10] considers an attacker whose coordinated attack is both cyber and physical in nature, and who selects the optimal time to launch attacks. The model is formulated over a 24-hour time horizon. In [11], an AD model is presented in which the defender's constraints include AC Optimal Power Flow (OPF) equations.

While the aforementioned studies have established their capabilities, there are also potential drawbacks:

1) Existing literature applies impact factors found from solving power flow equations to predict the targets of an attacker. This approach is not entirely valid; the decisions of attackers may well depend on cyber vulnerabilities in the communication infrastructure, and on the criticality of load.
2) By modeling the actions of defender and attacker as a single tri-level optimization problem, DAD models assume knowledge of the attacker's constraints by the defender. However, in practice, the attacker's constraints are unknown, and the true optimal defender action may be significantly different from what is found by the multi-level optimization problem. This assumption, thus, greatly reduces the practicality and usability of DAD models found in the literature. A preferred approach is for the defender to take optimal actions, having *learned* the motives of the attacker.
3) In both AD and DAD models, the attacker is assumed constrained by the number of lines/nodes they are able to attack. The assumption in DAD models, especially, is that a protected line/node cannot be attacked. Clearly, the assumptions may not be valid.
4) Some mitigation strategies proposed in the literature are empirical, the optimality of which is not proven. Furthermore, in both AD and DAD models, which seek to find an optimal mitigation strategy to an optimally launched attack, the mitigation strategy is implemented to reduce the impact of attacks, and not to prevent the attack. A preferred approach is to dynamically curtail the ability of the attacker before attack completion in an optimal manner.
5) Mitigation strategies found in the literature are either enforced at the communication level alone or the physical level alone. Mitigation on both levels is a holistic solution that provides greater resilience.
6) The existing literature determines coordination in attack and enforces mitigation from a centralized system. This can be prone to single point failures. A decentralized architecture is preferable.

B. CONTRIBUTIONS

This paper focuses on direct switching attacks on the electric power distribution system. The proposed algorithm consists of a decentralized system implemented by a multi-agent system (MAS), and a centralized system that performs network reconfiguration. The contributions of this paper are as follows:

1) A novel switching attack problem is formulated. Unlike the current literature, the attack model accounts for the influence of communication network vulnerabilities and criticality of load in the decisions of the attacker. Also, it models the attacker's constraints in terms of monetary costs. Therefore, the efficiency of any security mechanisms implemented a priori by the system defender and any cyber vulnerabilities impact the extent of the attacker's operations, given their budget. Furthermore, the model shows what type of attack is to be conducted, whether the attacker ought to take control over the switch, or if they ought to only deny the authorized user access to it.
2) A novel real-time decentralized mechanism for establishing coordination of attacks and predicting the targets of an attack is proposed. The technique, implemented by the MAS, uses machine learning to improve on predictions in a distributed fashion. It is not prone to single point failures. The coordination/prediction is performed without knowledge of the attacker's parameters. This makes it more practical than is found in the existing literature.
3) A hybrid mitigation strategy that combines both physical level and communication network level mitigation is proposed. Both levels of the strategy are optimized in real-time based on information learned from detected attacks. The physical level mitigation performs contingency analysis and distribution network reconfiguration (DNR). To the best of the authors' knowledge, this is the first combination of both physical- and communication-level mitigation, and which also leverages the learned behavior of the attacker.

The remainder of the paper is organized as follows. In Section II, a model of the distribution system, and of the attacker's behavior are presented. In Section III, the proposed two-part algorithm is summarized. The multi-agent part of the algorithm is discussed in-depth in Section IV, and the physical-level mitigation in Section V. In Section VI, simulations and their results thereof are discussed, and the paper is concluded in Section VII.

II. SYSTEM MODEL

A. ELECTRIC DISTRIBUTION SYSTEM MODEL

Electric distribution systems typically comprise a set of feeders/source nodes, $\mathcal{S}$, and their laterals. There is a set of demand nodes $\mathcal{D}$, and a set of lines, $\mathcal{E}$, that connects them to the feeders. Let $\mathcal{N} = \mathcal{S} \cup \mathcal{D}$. It must be noted that both generation, such as a distributed energy resource (DER), and load may be connected to the same node. For computational purposes, this case is considered as two nodes, with the load connected to one node (the demand node) and the generation connected to the other (the source node), and which are joined by a line of zero impedance. The demand nodes and feeders are connected through manual and remote-control switching and protection devices such as fuses and reclosers. Distribution systems are typically operated in a radial configuration, so that each node is connected to only one feeder at a time.

The operations center monitors the distribution network through a Supervisory Control And Data Acquisition (SCADA) system. The setup comprises a SCADA master at the operations center, and (feeder) remote terminal units (RTUs) that interface the remote-control switches. As far as SCADA and cyber issues are concerned, only remote-control switches are of interest. Let $\mathcal{N}_s$ be the set of nodes whose load are connected via remote-control switches, and $\mathcal{E}_s$ be the set of lines connected via remote-control switches.

Power flow equations characterize the voltage, current and power flow, and operation should be conducted within nodal voltage limits. Let $\overline{V}$ and $\underline{V}$ be the upper and lower voltage limits respectively. In a later section, the power system model is presented and used in optimal reconfiguration.

B. ATTACK OBJECTIVE MODEL

It is assumed that, in a distribution network, the goal of the attacker is to create islands, disrupting power supply to the load. This implies that the attacker may reconfigure the network to serve this purpose.

While there are various attack techniques, in this paper, it is assumed that the attacker is an external attacker who executes man-in-the-middle (MitM) attacks. These include replay, denial of service (DoS), packet modification/falsification, and password hacks. Consequently, only a limited amount of information is available to the attacker. Replay, packet falsification, and password hacks are useful for gaining access to toggle a switch. The choice of attack for toggling a switch is, therefore, dependent on the cost to attack and the likelihood of success. DoS, when performed alone, only denies the authorized user access to the switch. It is assumed that the attacker is aware of only the topology of the network, and the maximum size and composition of loads.

The attractive switches for the attacker are those that connect significant portions of critical load. This is measured by the attack quality', or simply, the quality' of the switch at node i, q_i, which is given by (1).

$$q_i = \kappa_i \sum_{\phi\in\Phi_i} P_{ic}{}^{\phi} \tag{1}$$

In (1), κ_i is the criticality of load at node i, while P_{ic}^{ϕ} is the current active demand on phase ϕ of node i prior to attack. It must be noted that the criticality indices are defined by the attacker. Thus, they may vary from attacker to attacker. In this paper, criticality indices for highly critical load (such as hospital load), critical load (such as industry load), and non-critical load are set at 2.0, 1.5 and 1.0 respectively. The criticality of the ith node, κ_i, is the weighted sum of criticality of its load components. For a switch that connects a line or generation, the attack quality is the maximum load quality lost when the switch is disconnected.

Moreover, some nodes may require fewer resources to attack than others. This may be the case if, for instance, a vulnerability is already found in firmware tools installed on the RTU deployed at the node. Furthermore, opening a switch may not be enough to execute a useful attack; the operator may serve load by closing other switches that are accessible to them. Therefore, in some cases, in addition to opening closed switches, the attacker must also keep an already open switch inaccessible to the operator. Hence, the attacker seeks to find the set of switches to toggle and the set of switches whose status must remain in the current state to maximize the total disrupted power. Since it is assumed that the actions of the attacker are not concealed from the operator, the attacker must maximize their reward subject to the operator maximizing the load served in the ensuing attack. Consequently, the attacker's behavior is modeled by a bi-level optimization problem as follows.

$$\begin{aligned}\max_{e^t,e^k,z}\ & R\Bigg[\sum_{i\in\mathcal{D},\phi\in\Phi_i}\kappa_i(P_{ic}{}^{\phi}-P_i^{\phi*})\Bigg]-\Bigg[C_F+\sum_{(i,j)\in\mathcal{E}_s}C_{e_t}^{ij}e_{ij}^t\\ &+\sum_{(i,j)\in\mathcal{E}_s}C_{e_k}^{ij}e_{ij}^k+\sum_{i\in\mathcal{D}_s}C_d^i(1-z_i)\Bigg]\end{aligned} \tag{2}$$

$$\text{subject to}\quad C_F+\sum_{(i,j)\in\mathcal{E}_s}C_{e_t}^{ij}e_{ij}^t+\sum_{(i,j)\in\mathcal{E}_s}C_{e_k}^{ij}e_{ij}^k+\sum_{i\in\mathcal{D}_s}C_d^i(1-z_i)\le B \tag{3}$$

$$e_{ij}^k \le 1-e_{ij}^t \quad \forall(i,j)\in\mathcal{E}_s \tag{4}$$

$$\sum_{i\in\mathcal{D},\phi\in\Phi_i}P_i^{\phi*}=\max_{P_g,P,y}\sum_{i\in\mathcal{D},\phi\in\Phi_i}P_i^{\phi} \tag{5}$$

$$P_{gi}^{\phi}=P_i^{\phi}+\sum_{j:(i,j)\in\mathcal{E},\phi\in\Phi_{ij}}P_{ij}^{\phi}-\sum_{j:(i,j)\in\mathcal{E},\phi\in\Phi_{ij}}P_{ji}^{\phi}\quad\forall i\in\mathcal{N},\phi\in\Phi_i \tag{6}$$

$$\sum_{j:(i,j)\in\mathcal{E},\phi\in\Phi_{ij}}P_{ji}^{\phi}=0\quad\forall i\in\mathcal{S},\phi\in\Phi_i \tag{7}$$

$$-s_{ij}F_{ij}\le P_{ij}^{\phi}\le s_{ij}F_{ij}\quad\forall(i,j)\in\mathcal{E}_s,\phi\in\Phi_{ij} \tag{8}$$

$$-F_{ij}\le P_{ij}^{\phi}\le F_{ij}\quad\forall(i,j)\in\mathcal{E}\backslash\mathcal{E}_s,\phi\in\Phi_{ij} \tag{9}$$

$$0\le P_i^{\phi}\le z_iP_{ic}^{\phi}\quad\forall i\in\mathcal{D}_s,\phi\in\Phi_i \tag{10}$$

$$0\le P_i^{\phi}\le P_{ic}^{\phi}\quad\forall i\in\mathcal{D}\backslash\mathcal{D}_s,\phi\in\Phi_i \tag{11}$$

$$y_{ij}\le 1-(e_{ij}^t+e_{ij}^k)\quad\forall(i,j)\in\mathcal{E}_s \tag{12}$$

$$s_{ij} = e^t_{ij}(1 - s^c_{ij}) + e^k_{ij}s^c_{ij} + y_{ij} \quad \forall (i,j) \in \mathcal{E}_s \tag{13}$$

$$e^t_{ij}, e^k_{ij}, y_{ij} \in \{0, 1\} \quad \forall (i,j) \in \mathcal{E}_s \tag{14}$$

$$z_i \in \{0, 1\} \quad \forall i \in \mathcal{D}_s \tag{15}$$

Objective function (2) maximizes the net reward of the attacker. Here, P^{ϕ}_i is the active load served at phase ϕ of node i, R is the monetary reward per kW power disrupted, e^t_{ij} and e^k_{ij} are binary variables that indicate whether a switch on a line is selected to be toggled or kept in its current state respectively, and z_i is a binary variable indicating the state of a switch connecting load at a node. The parameters C_F, and C^{ij}_{ek} are the fixed cost of attack, and the variable cost to keep a switch connecting line (i, j) in its current state, i.e., to perform DoS, respectively. For each switch the attacker has already determined which attack type to perform to toggle the switch. Thus, $C^{ij}_{e_t}$ and C^i_d are the variable cost to toggle a switch on line (i, j), and variable cost to toggle the switch connecting load at node i, respectively. Constraint (3) requires that the total cost of attack (i.e., sum of fixed cost and variable costs) is within their budget, B. Constraint (4) ensures that a disconnect switch selected to be toggled is not also selected to be kept in its current state.

Constraint (5) and the remaining constraints form the inner level optimization problem that characterize what the attacker believes will be the response of the operator to their attack. Constraint (5) maximizes the load served. In (6), the sum of power generated/supplied and the power inflow, at a node, equals the sum of power outflows and specified demand, and in (7) there is no power flow into a source node. The sets Φ_i and Φ_{ij} are the set of phases at node i, and the set of phases between nodes i and j, respectively. Variable P^{ϕ}_{ij} is the power flow on phase ϕ of the line between nodes i and j after attack. Line flow must be within acceptable limits in (8) and (9). However, if there is a remote-control switch on the line, (8) also constrains the line flow according to the state of that switch. Here, F_{ij} is the flow limit of the line between nodes i and j, and s_{ij} is a binary variable indicating the state of the remote-control switch on that line. Similarly, the served load at each demand node must be within its limits in (10) and (11), with (10) also constraining according to the state of the load remote-control switch. Constraint (12) implies that the operator can toggle only line switches that are not attacked. Thus, y_{ij} is the operator's parameter. Finally, the state of a remote-control switch on a line is determined by (13); it is dependent on whether the attacker selects it for attack, or what the operator sets it to be if the attacker does not select it for attack. In (13), s^c_{ij} is the current state of the remote-control switch on line (i, j). Equations (14)-(15) provide binary constraints. It is noted that radiality is not ensured here; the attacker has no need to ensure this constraint, and the operator may be unable to do so for a certain attack configuration. Since power flow is allowed to be negative, in this model, $(i, j) \in \mathcal{E}$ implies $(j, i) \notin \mathcal{E}$.

It is noteworthy that the bi-level mixed integer linear program (2) – (15) is computationally demanding, especially when solved for a large distribution system. For instance, when solved for the IEEE 123-Node Test Feeder, the computation takes more than 8 hours and the solution may not be exact. Thus, the method proposed in [12] is used to convert the problem into a single level mixed integer linear program: dual constraints of the inner level problem are added to the problem, and so is the equality constraint for the inner level dual and primal objectives.

III. SUMMARY OF PROPOSED ALGORITHM

The proposed algorithm consists of two parts: a decentralized system in the first part, and a centralized system in the second part.

The decentralized system is a multi-agent system (MAS), with an agent installed at each remote-control switch in the network. The agent is an autonomous software module, implemented on a computing device with RTU functionality. The agent is, therefore, the cyber interface of the remote-control switch, and is aware of attack quality, q, as well as its own operational properties. In this paper, the operational properties of an agent include the type of SCADA communication network it connects to (e.g., cellular network A, etc.), and the firmware. Each agent of the MAS executes a three-stage intrusion prevention algorithm, which includes attack detection, attack target prediction, and communication level mitigation.

On the other hand, the centralized system is located at the operations center and performs the second level of mitigation. It is activated when a user-defined K alerts have been received at the operations centers. An optimal reconfiguration of the network is performed to minimize potential loss of load. Following this, a command is issued by the central agent to dispersed agents to turn off remote control ability.

The proposed algorithm is hierarchical, enforcing a hybrid mitigation. A step-by-step summary is given in Fig. 1. The next sections detail the parts of the proposed algorithm.

IV. MULTI-AGENT (DECENTRALIZED) LEVEL

An agent is installed at each remote-control switch. Each agent implements the following three-stage algorithm.

A. DETECTION STAGE

In this first stage, the agent implements a network-based intrusion detection system (NIDS), which performs real-time checks in a sliding window of time, W. It monitors for a set of attacks $\{d, p, l, f\}$, where $d, p, l,$ and f are floo**d**ing, **p**assword login attempts, rep**l**ay and packet **f**alsification attacks, respectively. An event is classified as an attack if and when the number of anomalies within W exceeds pre-determined threshold values. In this paper, the Poisson distribution is used to determine such thresholds in an offline process; the arrival of communication packets, X, from the utility to a node is a Poisson process. A distribution is plotted for the period of W, and the minimum occurrence of x events such

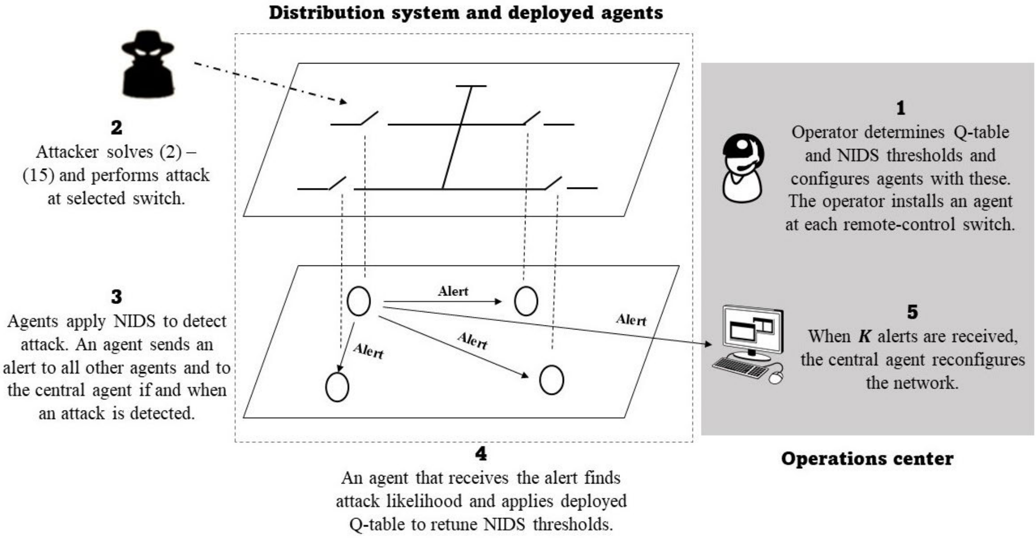

Distribution system and deployed agents
2
Attacker solves (2) – (15) and performs attack at selected switch.
3
Agents apply NIDS to detect attack. An agent sends an alert to all other agents and to the central agent if and when an attack is detected.
Alert
Alert
Alert
Alert
4
An agent that receives the alert finds attack likelihood and applies deployed Q-table to retune NIDS thresholds.
1
Operator determines Q-table and NIDS thresholds and configures agents with these. The operator installs an agent at each remote-control switch.
5
When K alerts are received, the central agent reconfigures the network.
Operations center

FIGURE 1. **Summary of proposed algorithm.**

that $Pr(X = x) \approx 0$ may be considered anomalous with respect to the utility's operations. Thus, x is the threshold, and the method is used to obtain thresholds for password login attempts, p_t, and flooding, f_t.

With the execution of packet authentication schemes, replay and packet falsification can be monitored. Since the utility sets up the necessary security keys for this, replayed and falsified packets are unexpected. Consequently, the threshold for replay, l_t and packet falsification, f_t may each be set to 1.

Should an intrusion be detected, the agent sends an alert, $\mathcal{L}$, to the operations center and to all other agents in the network. Denote the a**l**erting agent by n_L, and an agent that **r**eceives $\mathcal{L}$ by n_R. Among others, $\mathcal{L}$ contains the operational properties of agent n_L. All other agents activate the second stage of the decentralized algorithm when $\mathcal{L}$ is received. Meanwhile, n_L shuts down the remote control capability by blocking all incoming connections to itself and dropping all packets addressed to itself for a user-defined time period $T \leq W$. Only outgoing communication is allowed.

B. CORRELATION AND TARGET PREDICTION STAGE

Once an alert is received, agent n_R determines if its switch will be a target of the (ongoing) attack. Ideally, this requires an agent to solve the attacker's problem (2) – (15). However, agents are unable to do so since the reward, budget, and cost of attack are known only by the attacker. Consequently, a learning process for predicting the targets of an attack is used. The result is the attack likelihood index, ρ, which is a measure of similarity between n_L and n_R; it also measures the certainty of agent n_R that its switch will be a target.

It is assumed that nodes that possess similar operational properties have a similar cost to attack. An agent correlates its own properties to those in received alerts, establishing a measure of the cost to attack. The use of similarity is reasonable. Indeed, even in the case where the attacker randomly selects their targets, the existence of an entry point in the communication network will determine which targets are accessible to the attacker. Thus, the attacker's targets are still similar in that they operate a common vulnerable system.

Suppose that n_R has received a set of alerts within W, the sliding window of time for monitoring for attacks. The alerts were sent by different agents which detected attacks. Each alert contains the communication network type of the corresponding alerting agent, as well as its firmware type. Then, a set of unique communication network types in the alerts, $\mathcal{N}_L$, may be defined. Similarly, a set of the unique firmware types in the set of alerts, $\mathcal{F}_L$, may be defined. As more alerts are received, $\mathcal{N}_L$ and $\mathcal{F}_L$ are updated.

Meanwhile, the unique communication network types of all the agents in the distribution network form a set, $\mathcal{N}_N$, while the unique firmware types of all the agents form a set, $\mathcal{F}_N$. Then, $\mathcal{N}_L \subseteq \mathcal{N}_N$ and $\mathcal{F}_L \subseteq \mathcal{F}_N$. Agent n_R itself has a communication network type, $u^n_{n_R}$, and a firmware type, $u^f_{n_R}$. Define the following:

$$I_n := (u^n_{n_R} \in \mathcal{N}_L) \tag{16}$$

$$I_f := (u^f_{n_R} \in \mathcal{F}_L) \tag{17}$$

where $I(\cdot)$ is an indicator function. Definition (16) evaluates to 1 if the communication network of agent n_R is a member of the communication network types in the set of received alerts, and 0 otherwise. Similarly, (17) evaluates to 1 if the firmware type of agent n_R is a member of the firmware types in the set of received alerts, and 0 otherwise.

Variability in the operational properties in the set of received alerts signifies variability in the associated cost to attack. Subsequently, the metric of entropy is employed. Entropy, also known as Shannon's entropy [13], measures the degree of variability in a set; the lower the entropy, the more similar are the elements of a set, and vice versa. Entropy is a useful metric applied in machine learning and data mining programs such as decision trees to measure the quality of classification [14].

A similarity index for operational properties is calculated as follows:

$$\psi_{n_R} = I_n\left[I_f + H_f(1-I_f)\right] + H_n(1-I_n)\left[I_f + H_f(1-I_f)\right] \tag{18}$$

where H_n is Shannon's entropy of $\mathcal{N}_L$ relative to that of $\mathcal{N}_N$, and H_f is Shannon's entropy of $\mathcal{F}_L$ relative to that of $\mathcal{F}_N$. They are given in (19) and (20), respectively. Here, a and b are arbitrary elements of the given sets.

$$H_n = \frac{-\sum_{a\in\mathcal{N}_L} Pr(a)\log_2 Pr(a)}{-\sum_{b\in\mathcal{N}_N} Pr(b)\log_2 Pr(b)} \tag{19}$$

$$H_f = \frac{-\sum_{a\in\mathcal{F}_L} Pr(a)\log_2 Pr(a)}{-\sum_{b\in\mathcal{F}_N} Pr(b)\log_2 Pr(b)} \tag{20}$$

Since $\mathcal{N}_L, \mathcal{F}_L, \mathcal{N}_N$, and $\mathcal{F}_N$ contain only unique elements, it follows that for $a \in N_k$, $Pr(a) = \frac{1}{|\mathcal{N}_k|}$, for $k \in \{L, N\}$. Thus, (19) and (20) are simplified to (21) and (22) respectively.

$$H_n = \frac{\log_2 |\mathcal{N}_L|}{\log_2 |\mathcal{N}_N|} \tag{21}$$

$$H_n = \frac{\log_2 |\mathcal{F}_L|}{\log_2 |\mathcal{F}_N|} \tag{22}$$

From (21) and (22), it is clear that:

$$\lim_{|\mathcal{N}_L|\to|\mathcal{N}_N|} H_n = 1 \tag{23}$$

$$\lim_{|\mathcal{F}_L|\to|\mathcal{F}_N|} H_f = 1 \tag{24}$$

As $\log_2 |\mathcal{N}_N|$ and $\log_2 |\mathcal{F}_N|$ are constants, and the log function is monotonically increasing, H_n and H_f increase monotonically to 1. H_n and H_f represent the degree of randomness in the behavior of the attacker concerning their choice of communication network and firmware types. The following observations can be made from (18) - (24):

1) For $|\mathcal{N}_L| = 1$, $H_n = 0$. This indicates that the attacker has leverage from only one network type. A similar expression and interpretation can be derived for firmware type. When $H_n = H_f = 0$, only agents with the same properties as those in received alerts will have $\psi = 1$, whereas all others will have $\psi = 0$.
2) $|\mathcal{N}_L| = |\mathcal{N}_N|$ implies that $H_n = 1$ and $I_n = 1$ for all agents. Again, a similar expression can be written for firmware type. In this scenario, it is clear that the attacker has no preference for any communication network type, indicating that the firmware type and attack quality influence their choices.
3) $H_n = 0$ and $I_n = 0$ imply that $\psi = 0$, irrespective of the values of H_f and I_f. This is reasonable since an attacker must first gain access to a communication network before they are able to leverage any firmware vulnerabilities present.

Algorithm 1 Predicting the Likelihood of Attack

```
1:  if Load switch then
2:      if $z_i = 1$ then
3:          Continue
4:      else
5:          $\rho = 0$
6:          Break
7:      end if
8:  end if
9:  Determine normalized quality $q_i^n = \frac{q_i}{q_{max}}$
10: Determine correlation index $\psi_{n_R}$ for operational properties by (20).
11: $\mathbf{v_{n_R}} = [q_i^n \quad \psi_{n_R}]$
12: Determine worst case vector: $\mathbf{v_w} = \mathbf{1_2}$
13: $\rho = \frac{\sum_j \min\{v_{n_Rj}, v_{wj}\}}{\sum_i \max\{v_{n_Rj}, v_{wj}\}}$
```

The similarity index for operational properties found using (18) is used in Algorithm 1 to determine an attack likelihood index ρ_{n_R}, which measures the certainty of agent n_R of the likelihood of attack against its switch.

In Algorithm 1, the normalized quality, q_i^n, is used to ensure that the effect of ψ_{n_R} is not masked. Both q_i^n and ψ_{n_R} form a vector $\mathbf{v_{n_R}}$. It is noteworthy that $\mathbf{v_{n_R}} = \mathbf{1_2}$ (where $\mathbf{1_2}$ is the two-element one-vector) implies that agent n_R has the highest quality and the exact same operational properties as is present in received alert(s). This is the worst case scenario for the agent. Therefore, in step 6 of Algorithm 1, $\mathbf{v_{n_R}}$ is compared to the worst case scenario vector using the Jaccard similarity index [15], which performs an element-wise comparison of the two vectors.

C. COMMUNICATION-NETWORK-LEVEL MITIGATION

The attack likelihood index, ρ, found by n_R is a measure of the level of threat it perceives. As a response, agent n_R changes the thresholds used in monitoring for attacks, thus, reducing the likelihood of success of a potential attack. This makes the agent more attack-averse. Nonetheless, the new thresholds must not be overly restrictive to the authorized user; it must be commensurate with ρ. Q-Learning (QL), a reinforcement learning (RL) algorithm, may be used to determine the optimal new thresholds.

The output of QL is a Q-table which serves as a lookup table for the agent to determine how to change thresholds, given ρ. The change in thresholds is not dependent on the electrical operating conditions of the distribution network, but on ρ and the current value of thresholds, which form the QL state. Given the state, the optimal action is, therefore, how to change thresholds. The Q-table is obtained prior to the installation of agents in the distribution network and is deployed with agents. Therefore, it is not computed multiple times, but only once. The following subsection expounds the offline procedure of obtaining the Q-table from Q-Learning, while the next subsection explains how the table is used online once deployed with agents.

1) OFFLINE DERIVATION OF Q-TABLE

In this application of Q-Learning, the environment is the NIDS executed by the agent. It is desired that the increased level of protection offered by new thresholds matches the level of threat perceived, as measured by ρ. Therefore, the QL state, $s \in S$, is defined as the difference between ρ and the relative level of security offered by new thresholds. The state is given as in (25).

$$s = \rho - \left[1 - \frac{1}{4}\left(\frac{d_n}{d_t} + \frac{p_n}{p_t} + \frac{f_n}{f_t} + \frac{l_n}{l_t}\right)\right] \tag{25}$$

In (25), c_n represents new threshold values, while c_t represents the default threshold values, where $c \in \{d, p, l, f\}$. The thresholds, both default and new, for packet falsification and replay are each set to 1 for aforementioned reasons.

The state, s, is continuous and finite, ranging from -1 to 1, but is discretized in intervals of 0.05. The action set, A, contains eight actions: an increment action for each of d and p, a decrement action for each of d and p, a no-change action, set-to-default-thresholds action, set-to-median-thresholds action (i.e., $\{d_n, p_n, l_n, f_n\} \approx \{0.5d_t, 0.5p_t, l_t, f_t\}$), and shut-down-remote-control action (i.e., $\{d_n, p_n, l_n, f_n\} = \{0, 0, 0, 0\}$).

It is desired that new thresholds are selected to render $s = 0$. However, the discrete and finite nature of the thresholds implies that this not always achievable. Hence, some tolerance, τ, is allowed. A decaying exploration rate, ϵ, is used to balance exploration and exploitation. The reward in the next time step r_{t+1}, is as given in (26). Equation (26) gives a reward of 0 when the next state s_{t+1} after performing an action a_t is within a certain τ deviation from 0, and a penalty if it does not.

$$r_{t+1} = \begin{cases} 0 & |s_{t+1}| \leq \tau \\ -|\tau - |s_{t+1}|| & \text{otherwise} \end{cases} \tag{26}$$

Q-Learning is employed to find the optimal action value policy. As learning progresses, each action value $Q(s_t, a_t)$ in the Q-table is updated according to (27).

$$\begin{aligned} Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\Big[& r_{t+1} \\ & + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)\Big] \end{aligned} \tag{27}$$

The learning rate, α, is a hyper-parameter in the unit interval which determines how fast learning occurs. An α too small results in slow convergence. On the other hand, when selected too large, it leads to fast convergence to a potentially not optimal policy, or no convergence at all. To achieve convergence on a stationary environment as is the case in this application, a decaying α is used [16]. The discount factor, γ, also between 0 and 1, discounts future returns and ensures that the expected value is finite. The Q-table is updated according to (27) until the iteration converges.

2) ONLINE APPLICATION OF Q-TABLE

The Q-table obtained from the offline simulation is deployed with MAS agents, and a small ϵ is set to encourage exploration. When ρ is found in the second stage, an agent obtains its RL state using (25). The optimal action for that state is selected from the Q-table with a probability of $1 - \epsilon$. Upon applying the selected action, a new state is obtained. This process is repeated until the optimal action is to make no change to the thresholds. The final thresholds are enforced until T has elapsed, after which they are reset to their default values. The sliding time window W is also reset to start from when new thresholds are enforced. It is emphasized that the new thresholds are selected and enforced by n_R, which has not yet been attacked but is anticipating an attack with certainty ρ. This is, therefore, proactive communication level mitigation.

V. PHYSICAL-LEVEL MITIGATION

The decentralized algorithm implemented by the dispersed agents allows for attack target prediction and communication level mitigation. Suppose K alerts have been received at the operations center. It may be necessary to further perform a reconfiguration of the distribution system in order to minimize the impact of attacks. Contingency analysis is also useful to understand how many nodes and lines could be lost to the attacker without impacting distribution service. The central agent (CA) solves the following optimization problem.

$$\begin{aligned} \max \quad & \sum_{i\in\mathcal{D},\phi\in\Phi_i} P_i^\phi + \delta\Bigg(\sum_{(i,j)\in\mathcal{E}_s\setminus\mathcal{E}_{sp}} \rho_{ij}\Big[e^t_{ij}s^c_{ij} + e^k_{ij}(1 - s^c_{ij})\Big] \\ & + \sum_{i\in\mathcal{D}_s\setminus\mathcal{D}_{sp}} \rho_i(1 - z_i)\Bigg) \end{aligned} \tag{28}$$

$$\begin{aligned} P_{gi}^\phi = P_i^\phi + & \sum_{j:(i,j)\in\mathcal{E},\phi\in\Phi_{ij}} P_{ij}^\phi \\ & - \sum_{j:(i,j)\in\mathcal{E},\phi\in\Phi_{ij}} P_{ji}^\phi \quad \forall i \in \mathcal{N}, \phi \in \Phi_i \end{aligned} \tag{29}$$

$$\begin{aligned} Q_{gi}^\phi = Q_i^\phi + & \sum_{j:(i,j)\in\mathcal{E},\phi\in\Phi_{ij}} Q_{ij}^\phi \\ & - \sum_{j:(i,j)\in\mathcal{E},\phi\in\Phi_{ij}} Q_{ji}^\phi \quad \forall i \in \mathcal{N}, \phi \in \Phi_i \end{aligned} \tag{30}$$

$$\sum_{j:(i,j)\in\mathcal{E},\phi\in\Phi_{ij}} P_{ji}^\phi = 0 \quad \forall i \in \mathcal{S}, \phi \in \Phi_i \tag{31}$$

$$\sum_{j:(i,j)\in\mathcal{E},\phi\in\Phi_{ij}} Q_{ji}^\phi = 0 \quad \forall i \in \mathcal{S}, \phi \in \Phi_i \tag{32}$$

$$Q_i^\phi = P_i^\phi \tan\theta_i^\phi \quad \forall i \in \mathcal{D}, \phi \in \Phi_i \tag{33}$$

$$U_j^{\phi_j} = \sum_{i:(i,j)\in\mathcal{E},\phi\in\Phi_{ij}} w_{ij}\Bigg(U_i^\phi - 2\sum_{\phi\in\Phi_{ij}}\Big[r^{ij}_{\phi_j\phi}(\Gamma^{re}_{\phi_j\phi}P^\phi_{ij}$$

$$+\ \Gamma^{im}_{\phi_j\phi} Q^{\phi}_{ij}) + x^{ij}_{\phi_j\phi}(\Gamma^{re}_{\phi_j\phi} Q^{\phi}_{ij} - \Gamma^{im}_{\phi_j\phi} P^{\phi}_{ij})\Big]\Big]\Big)$$

$$\forall j \in \mathcal{N}, \phi_j \in \Phi_j \tag{34}$$

$$\underline{V}^2 \leq U^{\phi}_i \leq \overline{V}^2 \quad \forall i \in \mathcal{N}, \phi \in \Phi_i \tag{35}$$

$$0 \leq P^{\phi}_{ij} \leq w_{ij} F^{p}_{ij} \quad \forall (i,j) \in \mathcal{E}, \phi \in \Phi_{ij} \tag{36}$$

$$0 \leq Q^{\phi}_{ij} \leq w_{ij} F^{q}_{ij} \quad \forall (i,j) \in \mathcal{E}, \phi \in \Phi_{ij} \tag{37}$$

$$0 \leq P^{\phi}_{i} \leq z_i P^{\phi}_{ic} \quad \forall i \in \mathcal{D}_s, \phi \in \Phi_i \tag{38}$$

$$0 \leq P^{\phi}_{i} \leq P^{\phi}_{ic} \quad \forall i \in \mathcal{D} \backslash \mathcal{D}_s, \phi \in \Phi_i \tag{39}$$

$$0 \leq Q^{\phi}_{i} \leq z_i Q^{\phi}_{ic} \quad \forall i \in \mathcal{D}_s, \phi \in \Phi_i \tag{40}$$

$$0 \leq Q^{\phi}_{i} \leq Q^{\phi}_{ic} \quad \forall i \in \mathcal{D} \backslash \mathcal{D}_s, \phi \in \Phi_i \tag{41}$$

$$e^{k}_{ij} \leq 1 - e^{t}_{ij} \quad \forall (i,j) \in \mathcal{E}_s \tag{42}$$

$$s_{ij} = e^{t}_{ij}(1 - s^{c}_{ij}) + e^{k}_{ij} s^{c}_{ij} \quad \forall (i,j) \in \mathcal{E}_s \tag{43}$$

$$s_{ij} = s^{c}_{ij} \quad \forall (i,j) \in \mathcal{E}_{sp} \tag{44}$$

$$z_i = 1 \quad \forall i \in \mathcal{D}_{sp} \tag{45}$$

$$w_{ij} \geq 0 \quad \forall (i,j) \in \mathcal{E} \tag{46}$$

$$w_{ij} = 0 \quad \forall j \in \mathcal{S}, (i,j) \in \mathcal{E} \tag{47}$$

$$w_{ij} + w_{ji} = 1 \quad \forall (i,j) \in \mathcal{E} \backslash \mathcal{E}_s \tag{48}$$

$$w_{ij} + w_{ji} = s_{ij} \quad \forall (i,j) \in \mathcal{E}_s \tag{49}$$

$$\sum_{i:(i,j) \in \mathcal{E}} w_{ij} = 1 \quad \forall j \in \mathcal{D} \tag{50}$$

$$e^{t}_{ij}, e^{k}_{ij} \in \{0, 1\} \quad \forall (i,j) \in \mathcal{E}_s \tag{51}$$

$$z_i \in \{0, 1\} \quad \forall i \in \mathcal{D}_s \tag{52}$$

In (28), the CA, operating on behalf of the operator, seeks to find a new configuration that maximizes the total active power that can be served. However, the second term of (28) ensures that switches with high attack likelihoods are placed in the state desirable by the attacker, i.e., to keep open switches open and closed switches open. Hence, together, the two terms of (28) maximize the load served while minimizing the dependence on switches with high attack likelihood indices. In (28), δ is a scaling parameter. It has a unit of kW, and is chosen such that it is smaller than the magnitude of the smallest kW demand in the network. The expressions and parameters in (28)-(52) have similar interpretations as in (2) – (15). Constraint (33) relates active and reactive power using the load angle, θ. Constraint (34) is the squared voltage magnitude expression adapted from the linearized power flow model [17], which is constrained by (35) within the square of voltage magnitude limits. The linearized voltage equation, although an approximation, has been found to produce good results, giving an error of 1.06×10^{-3} pu voltage for a 2065-bus network [17]. In (34), r_{ij} and x_{ij} are the resistance and reactance matrices, respectively, of line (i, j), $\Gamma^{re}_{\phi_j\phi}$ is the element in row ϕ_j column ϕ of $\Gamma^{re} = Real\{\Gamma\}$. The interpretation is similar for $\Gamma^{im} = Im\{\Gamma\}$, where Γ is given by (53).

$$\Gamma = \begin{bmatrix} 1 & e^{-j4\pi/3} & e^{-j2\pi/3} \\ e^{-j2\pi/3} & 1 & e^{-j4\pi/3} \\ e^{-j4\pi/3} & e^{-j2\pi/3} & 1 \end{bmatrix} \tag{53}$$

TABLE 1. Operational properties assigned to agents.

Agent Location	Network Type	Firmware Type
Node 2	Network D	OS1
Node 24	Network D	OS4
Node 70	Network A	OS3
Node 88	Network B	OS2
Node 109	Network C	OS4
Line (451-450)	Network C	OS2
Line (300-350)	Network B	OS3
Line (251-250)	Network A	OS1
Line (151-300)	Network D	OS2
Line (150-149)	Network C	OS4
Line (97-197)	Network A	OS3
Line (61-610)	Network A	OS1
Line (60-160)	Network B	OS4
Line (195-95)	Network B	OS1
Line(18-135)	Network C	OS2
Line (13-152)	Network D	OS3
Line (54-94)	Network C	OS1

In (44) and (45), switches whose agents have shut down remote control capability remain in their current state. Here, $\mathcal{N}_{sP}$ and $\mathcal{E}_{sP}$ are the sets of node switches and line switches, respectively, whose agents have shut down remote control capability. The variable w_{ij}, adapted from [18] is introduced to determine the direction of power flow and ensure radiality in (46) – (50). Hence, in this model, $(i, j) \in \mathcal{E}$ implies $(j, i) \in \mathcal{E}$. Clearly, (28) – (53) is a non-linear integer programming problem due to the product of binary and continuous variables in (34). To ensure faster solution, (34) is linearized using big M notation.

VI. SIMULATIONS AND RESULTS

The IEEE 123-Node Test Feeder in Fig. 2, is used as a test system. The load at nodes 2, 24, 70, 88, and 109 are assumed connected via remote-control switches. Also, all three-phase line disconnect switches are assumed to be remote-control. Each load in the network is assigned a criticality level. Agents are implemented in Volttron [19], and Scapy, is used to create an attack platform. Fictitious operational properties, shown in Table 1, are assigned to each agent. The power system model is implemented in OpenDSS, and the single-level form of the attacker's problem and the linearized central agent's optimization problem are modeled using GAMS and solved using DICOPT and CPLEX solvers, respectively. Table 2 shows the variable attack costs for different experiments.

Simulations are grouped into three main subsections, VI-A - VI-C. In the first section, a direct switching attack is performed on a distribution system not implementing the proposed method. In the second section, offline planning stage simulations are performed. In the third section, the proposed algorithm is demonstrated on the IEEE 123-Node Test Feeder. In this paper, $K = 3$, and $T = W = 1$ hour. The attack reward, R, is 1 monetary unit (MU) per kW power disrupted.

A. RESPONDING TO DIRECT SWITCHING ATTACK WITHOUT PROPOSED ALGORITHM

In this simulation (E1), the proposed algorithm is not enforced. The goal is to restore as much load as is possible

TABLE 2. **Variable costs to attack switches (in MU) for different experiments.**

	E1		E2		E3		E4	
Line Switch	C_{e_t}	C_{e_k}	C_{e_t}	C_{e_k}	C_{e_t}	C_{e_k}	C_{e_t}	C_{e_k}
Line (451-450)	15.60	5.60	14.50	5.60	67.00	25.00	70.00	25.00
Line (300-350)	12.00	7.00	30.00	20.00	67.00	25.00	70.00	25.00
Line (251-250)	13.00	13.00	67.00	20.00	45.00	25.00	70.00	25.00
Line (151-300)	41.90	4.90	30.00	15.00	67.00	25.00	70.00	25.00
Line (150-149)	18.90	8.90	14.50	5.60	67.00	25.00	70.00	25.00
Line (97-197)	74.50	6.50	67.00	20.00	67.00	25.00	70.00	25.00
Line (61-610)	70.10	2.10	67.00	20.00	45.00	25.00	70.00	25.00
Line (60-160)	69.34	9.34	30.00	15.00	67.00	25.00	70.00	25.00
Line (195-95)	16.30	6.30	30.00	15.00	45.00	25.00	70.00	25.00
Line(18-135)	12.00	2.00	14.50	5.60	67.00	25.00	70.00	25.00
Line (13-152)	77.90	7.90	30.00	15.00	67.00	25.00	70.00	25.00
Line (54-94)	30.00	3.00	14.50	5.60	45.00	25.00	70.00	25.00
Load Switch	C_d		C_d		C_d		C_d	
2	30.45		30.00		45.00		70.00	
24	28.90		30.00		67.00		70.00	
70	35.78		67.00		67.00		70.00	
88	79.00		30.00		67.00		70.00	
109	58.21		14.50		67.00		70.00	

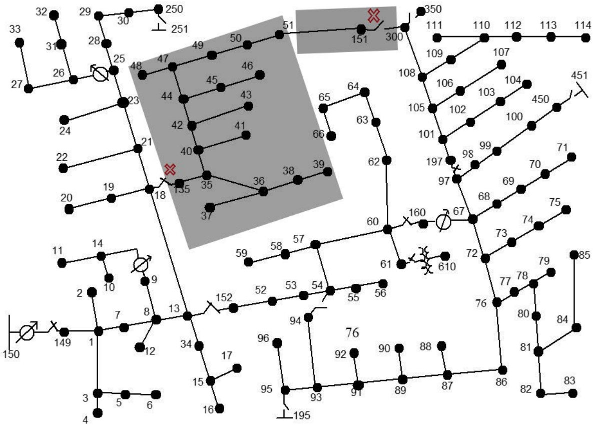


FIGURE 2. **Power outage caused by cyberattack.**

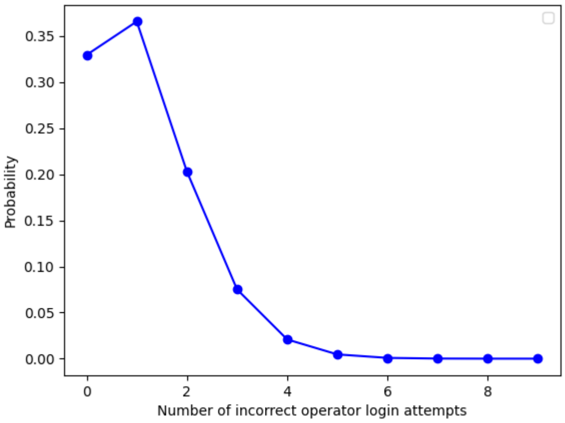


FIGURE 3. **A Poisson distribution of incorrect operator login events.**

after an attack. The attacker solves (2) – (15) in its single level form, using DICOPT solver. The computation takes 13 seconds, and the solution is exact. The attacker knows only the current state of switches, the network topology, and the current active demand being served. By using an aerial map of the location served by the target distribution system, the attacker infers the criticality of the load. All load switches are closed, and the states of all other disconnect switches are as shown in Fig. 2. The attacker's budget is 50 MU, and the fixed cost is 30 MU. For this configuration, the attacker finds that toggling the switch connecting line 18-135 and keeping the switch connecting line 151-300 in its current state, i.e., performing DoS, will yield a maximum net return of 1010.01 MU. The targets are shown with red crosses. This is a total of 755 kW of load. The shaded area in Fig. 2 shows the outage region. Since the attacked switches are inaccessible to the operator, the operator is unable to restore power to the outage region.

B. OFFLINE PLANNING STAGE SIMULATIONS

1) OBTAINING NIDS THRESHOLDS

Assume that from data collected at the operations center, the average incorrect login attempts by authorized operators is 1.27 within W. Using this average, the Poisson distribution plotted in Fig. 3 is obtained. The threshold is $p_t = 5$. This is the smallest value above which the probability of operator password error occurrence is 0. The average number of packets sent to a remote-control switch is 160. Since this mean is large, the Poisson distribution is approximated by the normal distribution. From this, the obtained threshold is $d_t = 200$.

2) OBTAINING Q-TABLE

There are forty discretized states. Increments and decrements to the threshold values are made in steps of 1 for each of login attempts and flooding. Certain combinations of thresholds have no logical interpretation (e.g., [0, 0, 1, 1]). Therefore,

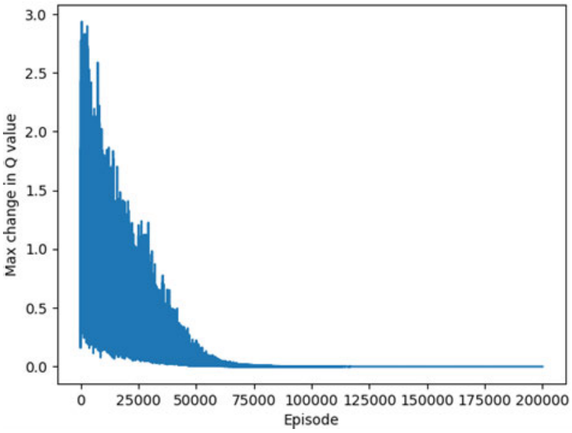


FIGURE 4. A plot showing the maximum absolute change in Q-value per episode.

a set of minimum values, $[d_n, p_n, l_n, f_n] = [2, 0, 1, 1]$ is set. For this combination, $p_n = 0$ is interpreted by the agent as not displaying the login console in its RTU interface. Next to this minimum set of thresholds is [0, 0, 0, 0], which is interpreted as shutting down remote control.

As a hyper-parameter, α is found from prior empirical tests, during which an initial value of 0.1 which decays slowly to 1×10^{-5} gives good results. Also, the discount factor, γ, is set to 0.99, while ϵ is set to 1 at the beginning of each episode and decayed to a minimum of 0.05. In addition, τ is taken as 0.05. Thousand trials are performed in each of two hundred thousand episodes. Convergence of the Q-Learning algorithm, shown in Fig. 4, occurs in 127 minutes. At convergence, a 6 KB Q-table is obtained which is deployed with agents configured with NIDS thresholds from the previous sub-section. The table is referenced during the communication level mitigation stage.

C. IMPLEMENTING THE PROPOSED ALGORITHM

Agents have been installed at all remote-control switches and the proposed algorithm is executed. Each agent has the NIDS thresholds and Q-table found from the previous sections.

1) LEVERAGE FROM COMMUNICATION NETWORK VULNERABILITY

Here (E2), all load switches are closed and the state of line switches is as shown in Fig. 5a. The attacker's budget is 100 MU, and the fixed cost is 50 MU. By solving (2) – (15), the attacker determines to toggle the switch connecting line 18-135 and load switch 109, and to create DoS at the switch connecting line 151-300. The anticipated disrupted load, shown in Fig. 5a as the green shaded area, is 795 kW which would give the attacker a reward of 1006.72 MU. They therefore begin a sequential attack, executing replay attack at switch 18-135, password hacks at load switch 109, and flooding switch 151-300. The agent at switch 18-135 detects the attack first, and sends an alert to both the CA and all other agents in the distribution network. All other agents calculate their attack likelihood indices and retune the NIDS parameters. As more alerts are received from the agents at switch 109 and switch 151-300, agents update the index and retune their NIDS thresholds with each update. For instance, by the third alert, the agent at switch 197-97 sets the following NIDS thresholds: [100, 2, 1, 1]. Fig. 6 shows the attack likelihood index, as updated by some agents as the alerts are received.

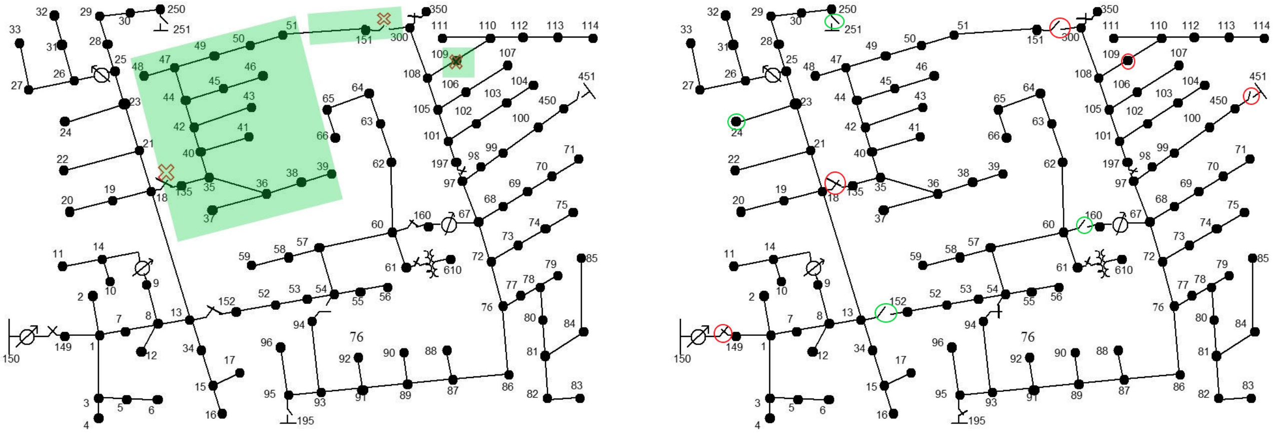
When $K = 3$ alerts are received by the operator, the central agent (CA) at the operations center queries the dispersed agents for their attack likelihood indices. The CA then solves (28) – (53) to determine a new configuration for the network, and follows with a command to turn off remote control capability for all switches. The new configuration, which serves all load, is shown in Fig. 5b. The red circles represent switches whose agents have shut down remote control capabilities, whereas the green circles represent switches with attack likelihoods greater than 0.4.

In Fig. 5b, four switches have attack likelihoods greater than 0.4. Of these, switch 251-250 remains in the same open state before and after the new configuration, while switches 13-152 and 60-160 are toggled (from closed to open) in the new configuration. By setting these new states, the CA has assumed that the three switches are in the control of the attacker and, thus, could be lost to the attacker. It is also observed that the new configuration uses feeder 195 to serve load as its switch has a relatively low attack likelihood. This is ideal; the central agent correctly uses switches with low attack likelihood indices to serve load, while avoiding reliance on those with high indices.

It is noteworthy that even though switch 451-450 is not targeted, its attack likelihood is high, as shown in Fig. 6, and it has shut down remote control capability. The switch has a high normalized quality, and its agent implements OS2 on Network C. When an alert is received from the agent at line switch 18-135 (which also implements OS2 on Network C), there is perfect operational similarity, i.e., $\psi_{451-450} = 1$. This illustrates proactive mitigation.

2) LEVERAGE FROM FIRMWARE VULNERABILITY

In this experiment (E3), OS1 has a firmware vulnerability. The attacker's budget is 150 MU, and there is a fixed cost of 50 MU. All load switches are closed, and the network is as shown in Fig. 7a. A firmware vulnerability is known but the attacker must first gain access to the network before leveraging this. By solving (2) – (15), the attacker determines to toggle switch 18-135 by sending a falsified packet, and flood switch 151-300. This coordinated attack, conducted simultaneously, is expected to yield a net return of 912.01 MU, being a disruption of 555 kW of load. Again, the agents deployed at the attacked switches detect the attack and send an alert to the CA and to all other agents in the distribution network. The alerting agents shut down remote control capabilities while the others retune their NIDS thresholds. However, $K = 3$ alerts are not received, and the CA is not triggered. No load is lost. The parameter, K, measures the willingness of the



(a) Anticipated outage area of attack

(b) New configuration from central agent (CA)

FIGURE 5. Results from experiment E2.

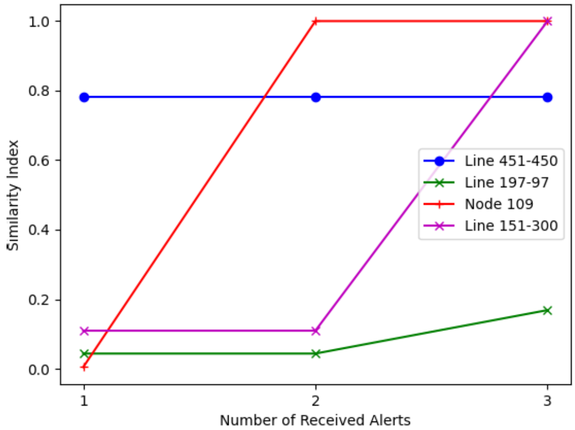
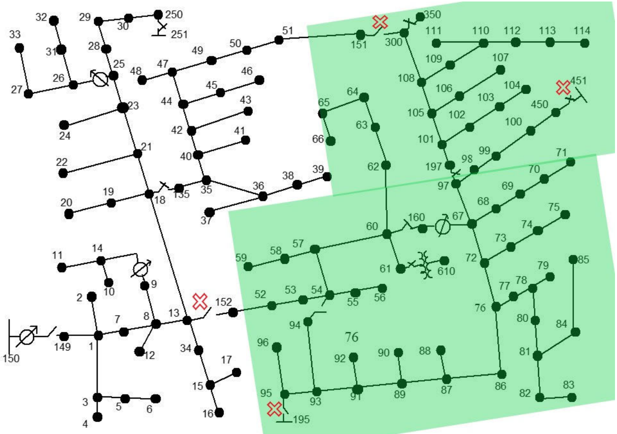
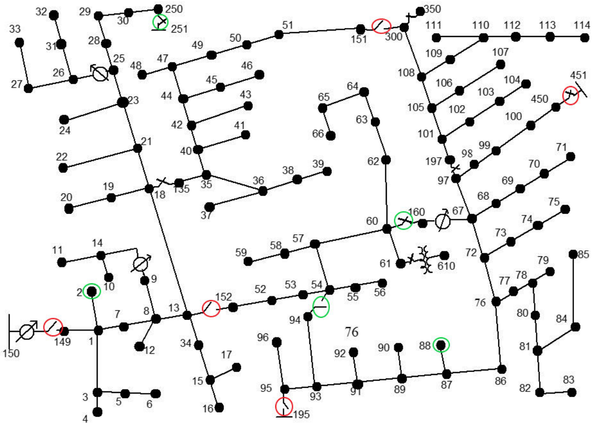


FIGURE 6. A plot showing change in attack likelihood, ρ, with receipt of alerts for some agents, in E2.

operator to reconfigure the network following the receipt of alerts. Thus, mitigation is left at the communication level for as long as is necessary.



(a) Anticipated outage area of attack



(b) New configuration from central agent (CA)

FIGURE 7. Results from experiment E4.

3) DEMONSTRATING THE IMPACT OF CRITICALITY

Here (E4), there is no known operational vulnerability. The attack cost is the same for all switches. The initial state of switches in this experiment is the same as for E3, shown in Fig. 7a. The attacker's budget is 200 MU and there is a fixed cost of 50 MU. Having solved (2) – (15), the attacker finds that they must toggle switch 451-450, and flood switches 13-152, 195-95 and 151-300. They anticipate a net return of 2748.59 MU, which is obtained from disrupting 1775 kW of load. This is shown in the green shaded area in Fig. 7a. A simultaneous coordinated cyberattack begins, and the agents deployed at the attacked switches detect the attacks. Although simultaneous, alerts are received in the following order: 451-450, 13-152, and 195-95. Before the central agent queries the dispersed agents for their attack likelihood indices, a fourth alert is received from the agent at switch 151-300. The dispersed agents update their attack likelihood index and retune their NIDS thresholds with each received alert. The central agent performs a new configuration, as shown Fig. 7b. The configuration is the same before and after the attack. This is mainly due to three of the feeder switches having shut down remote control capability. The CA immediately shuts down remote control capability for all remote-control switches still on line. No load is lost.

VII. CONCLUSION

This paper presents a decentralized attack correlation technique and a hybrid mitigation. Compared to interdiction models in the literature, this work assumes no explicit knowledge of the attacker's parameters by the defenders, which in this case, are agents. The targets of an attack are predicted in a decentralized manner using a learning mechanism, and new NIDS thresholds optimally found from reinforcement learning are applied. When enough alerts are received, physical mitigation is triggered. The proposed technique is also superior as it is not prone to single point failures; should the central agent be compromised, communication level mitigation is still enforced by the dispersed agents.

Currently, the NIDS implemented by the algorithm is anomaly-based and makes use of only communication level thresholds. It is therefore limited to only man-in-the-middle attacks. Future work may consider improving the mechanism of intrusion detection by integrating machine learning or another suitable method. Also, the inclusion of physical level checks in intrusion detection may prove useful for detecting insider attacks.

REFERENCES

[1] Electricity Information Sharing and Analysis Center (E-ISAC). (Mar. 2016). *Analysis of the Cyber Attack on the Ukrainian Power Grid, Electricity Information Sharing and Analysis Center (E-ISAC)*, [Online]. Available: https://nsarchive.gwu.edu/sites/default/files/documents/3891751/SANS-and-Electricity-Information-Sharing-and.pdf

[2] H. Zhang, B. Liu, and H. Wu, "Smart grid cyber-physical attack and defense: A review," *IEEE Access*, vol. 9, pp. 29641–29659, 2021.

[3] A. Gusrialdi and Z. Qu, "Smart grid security: Attacks and defenses," in *Smart Grid Control* (Power Electronics and Power Systems), 1st ed. Cham, Switzerland: Springer, 2018, pp. 199–223.

[4] R. Deng, P. Zhuang, and H. Liang, "False data injection attacks against state estimation in power distribution systems," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2871–2881, May 2019.

[5] S. Lakshminarayana, J. Ospina, and C. Konstantinou, "Load-altering attacks against power grids under COVID-19 low-inertia conditions," *IEEE Open Access J. Power Energy*, vol. 9, pp. 226–240, 2022.

[6] I.-S. Choi, J. Hong, and T.-W. Kim, "Multi-agent based cyber attack detection and mitigation for distribution automation system," *IEEE Access*, vol. 8, pp. 183495–183504, 2020.

[7] 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.

[8] C. Moya and J. Wang, "Developing correlation indices to identify coordinated cyber-attacks on power grids," *IET Cyber-Phys. Syst., Theory Appl.*, vol. 3, no. 4, pp. 178–186, Dec. 2018.

[9] Y. Lin and Z. Bie, "Tri-level optimal hardening plan for a resilient distribution system considering reconfiguration and DG islanding," *Appl. Energy*, vol. 210, pp. 1266–1279, Jan. 2018.

[10] K. Lai, M. Illindala, and K. Subramaniam, "A tri-level optimization model to mitigate coordinated attacks on electric power systems in a cyber-physical environment," *Appl. Energy*, vol. 235, pp. 204–218, Feb. 2019.

[11] A. Abedi, M. R. Hesamzadeh, and F. Romerio, "An ACOPF-based bilevel optimization approach for vulnerability assessment of a power system," *Int. J. Electr. Power Energy Syst.*, vol. 125, Feb. 2021, Art. no. 106455.

[12] A. L. Motto, J. M. Arroyo, and F. D. Galiana, "A mixed-integer LP procedure for the analysis of electric grid security under disruptive threat," *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1357–1365, Aug. 2005.

[13] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[14] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986.

[15] W. Wu, B. Li, L. Chen, C. Zhang, and P. S. Yu, "Improved consistent weighted sampling revisited," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 12, pp. 2332–2345, Dec. 2019.

[16] R. S. Sutton and A. G. Barto, "Multi-armed bandits," in *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018, ch. 2, pp. 32–33.

[17] L. Gan and S. H. Low, "Convex relaxations and linear approximation for optimal power flow in multiphase radial networks," in *Proc. Power Syst. Comput. Conf.*, Aug. 2014, pp. 1–9. [Online]. Available: https://ieeexplore.ieee.org/document/7038399

[18] J. A. Taylor and F. S. Hover, "Convex models of distribution system reconfiguration," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1407–1413, Aug. 2012.

[19] B. A. Akyol, J. N. Haack, S. Ciraci, B. J. Carpenter, M. Vlachopoulou, and C. W. Tews. (Jun. 2012). *Volttron: An Agent Execution Platform for the Electric Power System*. [Online]. Available: https://availabletechnologies.pnnl.gov/technology.asp?id=369

JENNIFER APPIAH-KUBI (Student Member, IEEE) received the B.S. degree in electrical engineering from the Kwame Nkrumah University of Science and Technology, Ghana, in 2016, and the M.S. and Ph.D. degrees from the Virginia Polytechnic Institute and State University, Blacksburg, VA, USA, in 2020 and 2022, respectively. She is currently a Cybersecurity Research Engineer with the Pacific Northwest National Laboratory, Richland, WA, USA. Her research interests include power grid analytics and renewable energy integration into the smart grid.

CHEN-CHING LIU (Life Fellow, IEEE) received the Ph.D. degree from the University of California at Berkeley, Berkeley, CA, USA. He is currently an American Electric Power Professor and the Director of the Power and Energy Center, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA. He is also an Adjunct Full Professor with University College Dublin, Ireland. He is an U.S. Member of the CIGRE Study Committee D2, Information Systems and Telecommunication, and a member of the U.S. National Academy of Engineering. He was a recipient of the IEEE PES Outstanding Power Engineering Educator Award, in 2004. He has served as the Chair of the IEEE PES Technical Committee on Power System Analysis, Computing, and Economics, from 2005 to 2006.

• • •