Version of Record: https://www.sciencedirect.com/science/article/pii/S0142061519344576
Manuscript_b25c1bc2669ba47a2884502cbcaacea3

# An Environment-adaptive Protection Scheme with Long-term Reward for Distribution Networks

Qiushi Cui, *Member, IEEE*, and Yang Weng, *Member, IEEE*

**Abstract**

Increasing renewable penetration in the distribution brings uncertainties, raising concerns for reliable grid operation. For example, relatively regular topological changes in distribution grids and the frequent on/off status of some distributed generators (DGs) add ambiguity to the short circuit levels of distribution networks. Consequently, protective relays need to adapt their settings to protect different operation conditions on distribution systems. Without such capability, relays may false trip or be insensitive. Previous methods ignore the long-term relay setting effect in relay coordination design. To bridge the gap, this paper proposes an environment-adaptive protection scheme (E-APS) to solve the protection coordination issue from a sequential decision making perspective. The agent-environment interaction is designed with protection knowledge integrated to enable the protection agent's adaptivity. After defining the state, action, and reward in reinforcement learning for the relay settings, we prove the convergence of the value function for post-decision state protection setting. In the numerical results, different system operation scenarios are applied to validate the performance of the proposed E-APS. This scheme is also compared with other optimization-based protection schemes. Results show that the E-APS is more adaptive to environmental change and achieves high performance in protection coordination.

**Index Terms**

Protection and control, renewable energy, reinforcement learning, R-GOOSE.

## I. Introduction

The power system is under increasing pressure from societal demands to become more secure, resilient, and efficient. Renewables are now cheaper than building new large-scale coal and gas plants. By 2050, the costs of an average PV and wind plant are expected to fall 71% and 58% respectively [1]. Batteries will further depress market prices, which in turn enable the deeper penetration of renewable energies like PV and wind. Though individually inverter-interfaced DGs do not affect fault currents much [2], the topological changes and number of DGs that will be interconnected to the grid will be capable

of significantly altering the fault current levels. The simple logic that commercial relays are typically programmed with [3], is incapable of making decisions in a changing environment and large state space brought on by the increased DG's.

To address the protection issues under altering system conditions, there are recent papers proposing setting-less protection solutions for modern distribution networks [4]–[7]. Setting-less protection has been inspired by differential protection, which has minimal settings and does not require coordination with other functions. This method processes the measurement data with a dynamic state estimation, which computes the best estimate of the protection zone states. However, the setting-less relays have a transitional issue when it comes to the parallel use of traditional and setting-less protection. For example, they have coordination problems with existing lateral protection means. Recently, the plug-and-play protection scheme is proposed [8], which can be categorized into the class of setting-less protection. This scheme is promising since it addresses the coordination issue with laterals and makes the protection scheme independent of a specific system. Although being flexible, the plug-and-play protection scheme requires a comparatively high bandwidth on its communication channels. Meanwhile, since this channel may be by power-line carrier (radio frequency over the power lines), audio tones over a telephone circuit or microwave channel, or a direct-wire fiber-optic pilot pair, high cost and high security to avoid undesired trip operation from extraneous signals on the channel are the major disadvantages [9]. Besides the setting-less protection, solutions for the protection and coordination issues under different system conditions come with three folders: (1) topological analysis, (2) optimization-based techniques, and (3) communication-based methods.

The relay setting issues can be handled through (1) proper topological analysis to decrease total fault clearing time. Examples include the conventional adaptive relay coordination scheme with the necessary control and communication medium in the distribution system [10]–[12], while improving existing relay settings in grid-connected and islanded distribution networks is suggested in [13]. To protect the radial and meshed feeders, dual setting directional overcurrent relay (DOCR) seems to be a suitable option in terms of less relay operating time compared to conventional relays [14]. However, coordination failure can happen when coordinating backup protection in the dual setting [15]. The dependence on the topological analysis and the limitation of conventional coordination philosophy jeopardises the adaptivity of the topology-analysis-based methods.

To improve the adaptivity of the DOCR performance, the coordination issue is generally examined through (2) optimization perspectives. For example, genetic-algorithm and the particle swarm optimization algorithm are used to achieve optimal coordination of DOCRs [16], [17]. The work in [18] employs an adaptive protection scheme to mitigate the DG impact on DOCR coordination using differential

evolution algorithm. In [19], the DOCR coordination problem is formulated as a mixed-integer non-linear programming problem and solved by the seeker algorithm. Moreover, the DOCRs coordination can be achieved through the nonlinear function and sequential quadratic programming method that is used for close and far-end fault location analysis [20]. However, the optimization-based methods drawbacks are determining the optimality of solutions and the omittance of the long-term relay effects.

The topology analysis, optimization, and communication based solutions belong to traditional machine learning control that has its own limitations since it relies on parameter identification or regression-based optimal actuation command computation. Their focus is the immediate cost function of the plant. However, the continuous change in system operational conditions alters the short-circuit levels. If the control on the relay settings does not consider the system-level variation, false tripping and other failures might occur. Therefore, an ideal "cost function" should consider the long-term reward of the relay settings. Such characteristics imply a continuous agent-environment interaction that leads to the adoption of reinforcement learning. The original contributions of this paper include:

- Designing an adaptive protection scheme by incorporating the inverse-time current characteristics into the reinforcement learning setup with post-decision states. To the best of our knowledge, this paper is the first work that introduces reinforcement learning to DOCR coordination. We design the state, action, and reward for an adaptive protection scheme that leverages on reinforcement learning. The post-decision states are introduced to simplify the problem's complexity.
- Development of an environment-driven protection coordination scheme in distribution systems deploying Routable Generic Object Oriented Substation Event (R-GOOSE). The concept of void action is introduced in the protection field to capture its uniqueness against general reinforcement learning problems. Meanwhile, a viable solution of applying the peer-to-peer protocol of R-GOOSE is proposed on top of the reinforcement learning algorithm.
- A convergence-guaranteed value iteration method and a Q-learning design for relay coordination, which enable the relay settings to be adaptive to the environment. We mathematically prove the convergence of the relay setting's value function in post-decision states and the value-iteration method. Furthermore, the Q-learning algorithm for relay setting coordination is proposed to achieve the adaptive relay setting.

Paper outline is as follows: Section II demonstrates the required communication hierarchy for the adaptive protection scheme. Section III provides the details of the reinforcement learning design and its theoretical support. The numerical results are shown in Section IV. Section VI is the Conclusions.

## II. Communication Hierarchy

Before introducing the proposed protection scheme, we first explain the communication hierarchy that the proposed E-APS deploys. Generic Object Oriented Substation Event (GOOSE) message is widely used in protection and distribution automation. The GOOSE message was originally designed to be used within a substation. To utilize such messages over wide-area networks under modern cyber security requirements, IEC TC 57 provides routable profiles for IEC 61850-8-1 GOOSE (called R-GOOSE), the packets of which can be utilized to transport general IEC 61850 data as well as synchrophasor data [21]. The communications are based on the full seven-layer Open System Interconnection stack and use Transmission Control Protocol (TCP)/User Datagram Protocol (UDP) multicast. R-GOOSE applications [22], [23] are appearing in wide-area protection and control field in recent years as the evolution of distribution automation standard – IEC 61850. With the system monitoring function enabled by R-GOOSE [24], we can detect:

- a change in the electric power system topology,
- a change in system load and generation,
- and the on/off status of DGs.

In this paper, we take full advantage of R-GOOSE messages. As shown in Fig. 1, the adopted hierarchy relies on GOOSE messages that support peer-to-peer communications among intelligent electronic devices (IEDs) with flexibility and satisfies the high-speed communication requirement in distribution networks. The R-GOOSE repetition mechanism ensures that the message is sent with a changing time interval between the repeated messages until a new change event occurs [25] and provides a secure communication channel between the protective relays.

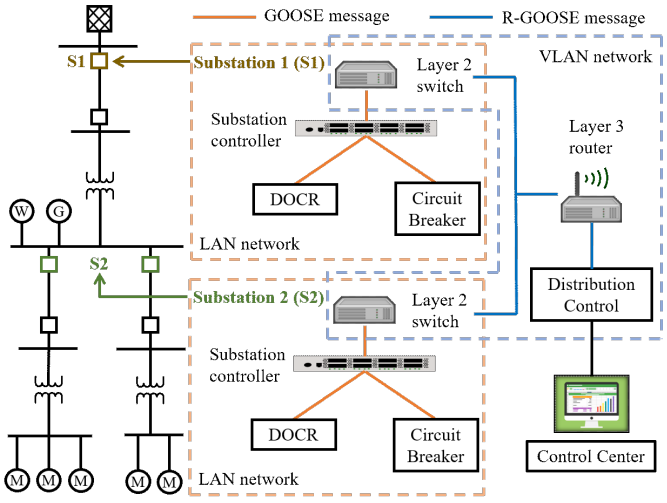



Figure 1. Hierarchical distribution automation system.

Implemented in the control center, the E-APS relies on the existing SCADA system; therefore it does not require extra infrastructure investment. The E-APS monitors the topology and operational conditions by processing data transmitted from Layer 2 and 3 switches. The data includes the on/off status of circuit breakers and DGs, system load/generation, and number of DGs in operation. Meanwhile, load and generation forecast data are also acquired by the control center. By comparing the data requirement between the E-APS and other existing methods [14]–[18], [26] in literature, the E-APS does not require extra data measurements. However, it is noteworthy that the E-APS requires the modeling of the distribution feeder in power system software that is capable of short-circuit analysis and has programming language API (Application Programming Interface). Such functions are available in most of the off-the-shelf power system software like Etap, CYME. This is the trade-off to achieve intelligence by interactively learning the system response before making the optimal decision in the dynamic conditions. Comparing with other existing methods, the extra work required in the E-APS is not the data but the learning agent design and its interaction with the distribution model. More details will be discussed in the next section.

Once a change in the system is identified, the E-APS conducts load flow, short circuit, contingency, and sensitivity analysis, then sends the environment data to each DOCR through R-GOOSE. Among different DOCRs, communication is mainly required from downstream DOCR to its adjacent upstream DOCR in radial networks, containing only the instantaneous response time. To achieve fast relay response and distributed computation, the learning model of each DOCR is stored locally.

The advantages of low installation cost, sufficient data rates, and ease of deployment make the industrial wireless local area network (WLAN) technologies popular among power utilities, especially for less critical smart distribution network applications. The proposed E-APS utilizes WLAN for data exchange, relying on the electromagnetic transfer of information at the edge and Ethernet cable as the bulk network. Recent studies have demonstrated the suitability of WLAN technologies for industrial environments. The WLAN technologies offer several data rates, such as $1$, $11$, and $51\,\mathrm{Mbps}$. Taking the above three data rates into consideration, the average GOOSE message delay is between $3.8$ and $4.2\,\mathrm{ms}$, whereas the maximum delay ranges from $4.2$ to $15.2\,\mathrm{ms}$ [27]. This delay is insignificant for the proposed method since the protection of AC distribution networks is less time-critical and not significant when compared with DC networks and transmission networks. Additionally, the backup is carefully addressed in the proposed method. Since this paper focuses on the topological changes and the on/off status of DGs, the delay of around several tens of milliseconds is tolerable for this overcurrent protection scheme, which is evidenced by [27].

## III. Intelligent Adaptive Protection Scheme

Environmental adaptivity is important to relay coordination schemes since the topological change and DG on/off status alter the short-circuit levels frequently. If the short-circuit levels change but the protective relay settings remain unchanged, the relay coordination scheme is at risk of false tripping and other failures. To resolve this issue, we are motivated to design an interactive and goal-seeking agent that optimally responds to a stochastic environment. This will deviate from the previous optimization-based solutions by considering the maximum return that quantifies its long-term objective. Through investigation, we adopt reinforcement learning (RL). RL is widely used in in demand response, energy management, cybersecurity, and power system control. Unlike supervised learning that learns the deterministic actions from labeled datasets or unsupervised learning that unveils the hidden structure from unlabeled datasets, RL guides its own experience acquisition from a stochastic system by maximizing its reward values in an interactive way. The RL-based control algorithm can be applied with certain system uncertainties. Contrarily, traditional control or optimization algorithms cannot handle stochastic scenarios effectively.

The environment of a power system protection set up includes the electric grid status – the short-circuit level and the protective device setting. We define two types of actions taken by the protection agent: one is the direct action $a_t$ that has a direct impact on the protection environment; we call the other one the void action $a_\emptyset$ since it hardly affects the environment. Although the system control can be involved to affect the short-circuit level, we keep the design of void action since it is critical to isolate protection functions from control functions and guarantee the reliability of the protection scheme. The agent-environment interaction in protection setup is plotted in Fig. 2. The protection agent – usually a distribution system operator (DSO) – learns the optimal actions to be taken towards a long term benefit.

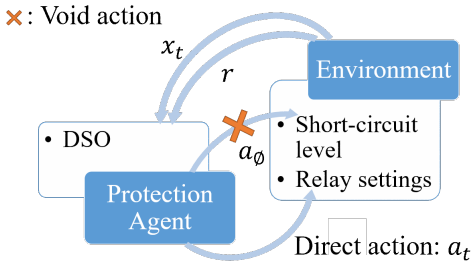



Figure 2. Agent-environment interaction in protection setup.

### A. Design of the State, Action, and Reward

We carefully design the three primary elements in reinforcement learning by embedding our power engineering knowledge into the learning procedure. Generally speaking, the state is our measurement, the action is the relay setting we choose, and the reward is the measured performance changes.

*1) State:* The *state* stems from the observation of the environment. Concretely, it contains short-circuit currents at each bus in the network, as well as the relay settings. The relay settings in this paper include the time dial settings (TDSs) and the pickup currents. The introduction of the environment-related state enables the interaction of the protective devices with the power networks. In modern digital relays, the TDS and pickup current are both continuous states. Certainly, the short-circuit current is also a continuous state of the environment. Assuming there are $N$ DOCRs, we can define a 3-tuple state for the $i$th DOCR:

$$x_{i\in N} = (x_{relay}, x_{net}) = ((TDS_i, I_{pu,i}), I_{sc,i}). \tag{1}$$

*2) Action:* The action involves increasing and decreasing the TDS and pickup current. It is noteworthy that the *action* does not affect the power networks ($x_{net}$) but the state of relays ($x_{relay}$). We, therefore, assign zero to the environment-related action. Similar to the operation of a knob of a thermostat, the *action* is with both positive and negative values. Since the state is continuous, the associated action is also continuous:

$$a_{i\in N} = (a_{relay}, a_{\emptyset}) = ((a_{TDS_i}, a_{I_{pu,i}}), 0). \tag{2}$$

*3) Reward:* The reward is based on the environment variation as well as the goodness of relay coordination setting. It is sensitive to the line topology, load connectivity, DG connectivity, the sum of all overcurrent relays' tripping time, the number of coordination violation, and the coordination time interval (CTI). It is defined as follows:

$$r_{i\in N} = \begin{cases} -T_{op,i}, & T_{op} \geq (T_{i-1} + \text{CTI}_i), \\ -bT_{op,i}, & \text{otherwise}, \end{cases} \tag{3}$$

where $b$ is the coordination penalty ratio and $b > 1$ in most of the cases to punish the improper coordination settings. At the fault location of $j$, $T_{op,i}$, the operating time of the $i$th DOCR in seconds, is expressed as

$$T_{op,i\in N} = \text{TDS}_i \cdot \left(\frac{A}{M^p - 1} + B\right), \tag{4}$$

where $\text{TDS}_i$ is the time dial setting of the $i$th DOCR, $A, B, p$ are constants chosen to provide the selected curve characteristics, $M = I_{\text{input},ij}/I_{\text{pickup},i}$, and $I_{\text{pickup},i}$ is the relay current set point. The variable of $I_{\text{input},ij}$ is determined by the network status such as the topology, the generation, and load profiles.

### *B. Protection Scheme Relying on Reinforcement Learning*

The assumption on the protective relay's terminal state is firstly introduced. According to relays' distinctive characteristics, we utilize the post-decision states to simplify the learning procedure. Next, we explain the value iteration algorithm and prove its convergence under post-decision states. Then, a Q-learning algorithm designed for relay setting coordination is elaborated.

**Assumption 1.** *(**Termination State of Protective Relays**) For protective relays under any setting policy, there exists a positive integer $d$, the probability of not reaching the terminal state $x_{ter}$ after $d$ steps is less than $1$ regardless of the initial state $x_0$:*

$$\rho_\pi = \max_{i=1,\cdots,n} P\{x_d \neq x_{ter} | x_0 = i, \pi\} < 1. \tag{5}$$

Potential terminal states can be the scheduled maintenance or the forced state of the relays.

*1) Post-decision states:* Through our investigation, we realize that the number of actions is large, usually twice the size of state space[1]. To avoid the corresponding complexity, we adopt the post-decision states. Given the setup of the protection devices, a set of post-decision states $Z$ has a smaller size than all the possible state-action pairs. Mathematically, the transition probability is expressed as

$$\mathcal{P}(x, a, y) = \mathcal{P}_A(f(x, a), y), \ x, y \in \mathcal{X}, a \in \mathcal{A}. \tag{6}$$

where the function $f$ represents the deterministic effect of the actions, whereas $\mathcal{P}_A$ captures the stochastic effect.

The post-decision state makes the learning of the immediate reward function easy. It is more economical and efficient than learning an action-value function in the protection setup. We define the post-decision sate optimal value function below:

$$V_A^*(z) = \sum_{y \in \mathcal{X}} \mathcal{P}_A(z, y) V^*(y), \ z \in Z. \tag{7}$$

Consequently, the update rules and decision strategies can be based on the following value function:

$$Q^*(x, a) = r(x, a) + \gamma V_A^*(f(x, a)), \tag{8}$$

where $Q^*(x, a)$ is the optimal action-value function, $r(x, a)$ is the reward given state $x$ and action $a$, and $\gamma$ is the discount factor.

*2) Value iteration (VI) algorithm:* In the post-decision state space setup, we design the VI algorithm for the protection coordination problem with post-decision states as follows:

[1]For example, considering only the TDS with a state space from 1 to 15, we need actions ranging from $-14$ to 14 to bring the current state to any state in the state space. However, with post-decision states, we only need to consider 15 states.

**Algorithm 1:** VI for Post-decision State Protection Setting

**Result:** output a deterministic policy, $\pi \approx \pi_*$, such that $\pi(x) = \arg\max_a \sum_{y,r} \mathcal{P}_A(f(x,a),y)[r(x,a) + \gamma V(y)]$.

Initialize a small threshold $\theta > 0$ indicating estimation accuracy;

Initialize $V(x)$, for all $x \in \mathcal{X}$, arbitrarily except that $V(\text{terminal}) = 0$;

**while** $\Delta > \theta$ **do**

  $\Delta \leftarrow 0$ ;

  **for** *each* $x \in \mathcal{X}$ **do**

    $v \leftarrow V(x)$;

    $V(x) \leftarrow \max_a \sum_{y,r} \mathcal{P}_A(f(x,a),y)[r + \gamma V(y)]$;

    $\Delta \leftarrow \max(\Delta, |v - V(x)|)$;

  **end**

**end**

In value table $V(x)$, we deploy the structure as shown in Fig. 3. In this application, $x_{relay}$ and $x_{net}$ include only $I_{pu}$, TDS, and $I_{sc}$. More elements can be added in the future. We assume there are $n$ $I_{pu}$ states, $m$ TDS states, and $k$ $I_{sc}$ states. Fig. 3 reveals each TDS cell contains $n$ $I_{pu}$ states and each $I_{sc}$ cell contains $m$ TDS arrays.

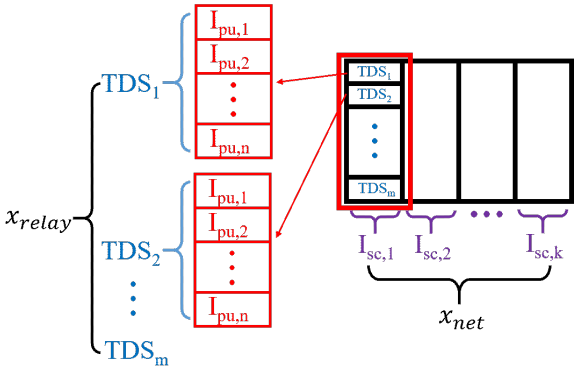



Figure 3. Value table structure.

**Theorem 1.** ***(Convergence of VI with Post-decision States)*** *Fix a deterministic policy $\pi \in \Pi$, $x, y \in \mathcal{X}, a \in \mathcal{A}$, an optimal value function defined from the Bellman optimality equation*

$$V^*(x) = \max_a \sum_y \mathcal{P}_A(f(x,a),y)[r + \gamma V^*(y)], \tag{9}$$

*converges to optimal value $V^*$.*

*Proof.* For any estimate of the value function $\hat{V}$, we define the Bellman operator $B$: $\mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$, by

$$B\hat{V}(x) = \max_a \sum_y \mathcal{P}_A(f(x,a),y)[r + \gamma \hat{V}(y)].$$

Actually, the Bellman operator $B$ follows the contraction mapping theorem: given any value functions $V_1, V_2$, we have

$$|BV_1(x) - BV_2(x)| \leq \gamma ||V_1 - V_2||_\infty \tag{10}$$

where $||\cdot||_\infty$ denotes the infinity norm. Due to page limit, the proof of equation (10) is provided in Appendix A. Consequently, with (10) we have

$$\begin{aligned} ||V_{k+1} - V^*||_\infty = ||BV_k - BV^*||_\infty \leq \gamma ||V_k - V^*||_\infty \\ \leq \cdots \leq \gamma^{k+1} ||V_0 - V^*||_\infty \end{aligned}$$

Since $0 \leq \gamma < 1$, as $k \to \infty$, the value function $V_k$ converges to the optimal value $V^*$.

□

*3) Q-learning and its convergence:* Traditional machine learning control requires an explicit system model and the cost structure. However, such a model is difficult to obtain under complex system operation conditions. Thus, we deploy Q-learning to overcome these issues. Utilizing value iteration, it can be directly used in the case of multiple policies. Since Q-learning directly updates the estimates of the Q-factors associated with an optimal policy, no complicated policy evaluation steps are required. Such a characteristic is suitable for the 3-tuple relay state space.

Given any initial estimate of the state-action value $Q_0$, action $a, b \in \mathcal{A}$, Q-learning adopts the following update rule:

$$\begin{aligned} Q_{t+1}(x_t, a_t) = Q_t(x_t, a_t) + \alpha_t(x_t, a_t) \\ [r_t + \gamma \max_b Q_t(x_{t+1}, b) - Q_t(x_t, a_t)], \end{aligned} \tag{11}$$

where the step size $\alpha_t$ rests in the range of $[0, 1]$.

**Theorem 2.** ***(Convergence of Q-learning)*** *With the update rule in equation (11), $Q_t$ converges with probability 1 to the optimal Q-function as long as*

$$\sum_t \alpha_t(x,a) = \infty, \quad \sum_t \alpha_t^2(x,a) < \infty, \tag{12}$$

for $0 \leq \alpha_t < 1$, and all $(x, a) \in \mathcal{X} \times \mathcal{A}$.

**Algorithm 2:** Q-learning for Relay Setting Coordination.

**Result:** output the action-value function estimate.

Initialize $Q : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$ arbitrarily, $x, y \in \mathcal{X}$, $a, b \in \mathcal{A}$;

Initialize a small threshold $\epsilon > 0$ indicating convergence tolerance;

**while** $|\alpha \cdot \delta| > \epsilon$ *and* $x$ *is not terminal* **do**

  **for** *each post-decision state* $x \in \mathcal{X}$ **do**

    $\pi(x) \leftarrow \arg\max_a Q(f(x, a))$;

    $a \leftarrow \pi(x)$;

    $r \leftarrow r(x, a)$;

    $y \leftarrow f(x, a)$;

    $\delta \leftarrow r + \gamma \max_b Q(y, b) - Q(f(x, a))$;

    $Q(f(x, a)) \leftarrow Q(f(x, a)) + \alpha \cdot \delta$ ;

    $x \leftarrow y$;

  **end**

**end**

The proof of Theorem 2 is provided by [28]. To achieve requirements in (12), we design Algorithm 2 to ensure that all state-action pairs are visited sufficiently often[2].

### *C. Complexity Analysis of the E-APS*

The complexity analysis and computational cost of online reinforcement learning have been studied and developed by researchers in the 1990s [29], [30]. It has been shown that the task of reaching a goal state for the first time is $O(n^3)$ if there are no duplicate actions [29]. This worst-case complexity provides an upper bound on the complexity of the E-APS. Since the E-APS uses only the Q-values, it performs undirected exploration. The upper bound indicates that the complexity under undirected exploration is polynomial in $n$ – the number of states. Furthermore, according to [29], the complexity of the E-APS can be decreased with suitable initial Q-values that are calculated offline. We have already formulated a good task representation in Section III-A with value iteration as discussed in Section III-B2. These domains, including the state and action space, have additional properties. The state space topology in the E-APS has a linear upper action bound $b \in \mathcal{N}_0$ iff $e \leq bn$ for all $n \in \mathcal{N}_0$. Then, the worst-case complexity becomes $O(bn^2) = O(n^2)$. Meanwhile, the determination of initial Q-values, which is a large portion of calculation, can be efficiently conducted offline, resulting in much faster convergence and less

[2]Due to the computational resource limitation, the relay setting state-action space cannot be visited for infinite times. However, we can visit this space for finite times until the incremental of Q value falls below a threshold.

complexity. This proposed E-APS is tractable with a polynomial order of complexity and therefore able to scale up to handle real-world problems in real-time.

### *D. Exploration in Large State Space to Guarantee Adaptivity*

Distribution networks are dynamic, therefore, besides the topological and operation condition alteration, we consider the following aspects: the short-circuit level increase due to the changes in the transmission system, the penetration increase of DGs, and a large number of outage and contingency cases reported from our utility partner. We utilize the $\epsilon$-greedy method for balancing exploration/exploitation. For such an application, a small $\epsilon$ value is recommended based on an extensive case study. Since the proposed RL method needs to behave non-optimally to explore all states and actions, it requires various short circuit simulations to generate behavior. In statistics, importance sampling is a general technique for estimating expected values under one distribution given samples from another [31]. This technique is suggested to train the protection agent to find the optimal actions "off" the target policy – a policy that is based on 24-hour realistic environment data. Instead, more short circuit scenarios are simulated to learn a general behavior by increasing the short circuit probability. More details can be found in [32] and we will not elaborate on it since it is not the focus of this paper.

### *E. Coordination with Reclosers and Fuses*

The E-APS can be seamlessly coordinated with other protective means on distribution feeders with laterals, where (1) reclosers are used at the beginning or midpoint of the feeder in coordination with its downstream DOCRs and (2) fuses are used on the laterals in fuse-saving or fuse-blowing strategy. In protection systems, reclosers save the electric companies considerable time and expense, since they permit power to be restored automatically, after several attempts. Meanwhile, a fuse is an automatic means of removing power from a faulty system at a low cost. Therefore, it is significant to design an E-APS that coordinates with the reclosers and fuses to minimize their effect. Fortunately, the proposed E-APS mainly handles the coordination among all DOCRs, therefore, we can deal with the coordination of E-APS DOCR with reclosers and fuses in a conventional way. The DOCRs under E-APS coordinate with all downstream devices whether fuses or reclosers by the desired CTI (for example, we can set the relay–fuse total clear: 0.2 sec, relay–series trip recloser: 0.4 sec, relay–relayed line recloser: 0.3 [9], [33]). In the E-APS, these CTIs are the prior knowledge that is included in equation (3). In such a way, the reclosers and fuses are well coordinated with the E-APS without jeopardizing it.

## IV. Numerical Results

The proposed E-APS method is tested extensively under multiple systems. In this section, we briefly explain two benchmark systems and then demonstrate the results. The results on both networks are presented from two distinct perspectives. In the smaller system we show results from the reinforcement learning perspective, including the Q-table, action table, and iteration numbers. The larger system reveals the results at the system level, including the overall system setting, environment-adaptive setting adjustment, and the comparison with other methods.

### *A. Generalized Systems Under Study*

The proposed E-APS is demonstrated in two representative systems. The first one is a modified 14-bus radial test feeder from CYME, shown in Fig. 4. It has 7 relay coordination pairs, a wind farm (W), two synchronous generators (G), ten motors (M) and a spot load. The second system, also with $13.8\,\mathrm{kV}$ voltage level, is a modified 38-bus system from CYME. Its single-line diagram is shown in Fig. 5 with 18 coordination pairs. The loads include 27 motor loads and 2 spot loads, summing up to $32\,\mathrm{MVA}$. DGs in the feeder contain three synchronous generators (9, 12.5, and $15.6\,\mathrm{MVA}$) and fourteen PV systems of $20.18\,\mathrm{MW}$ capacity in total.

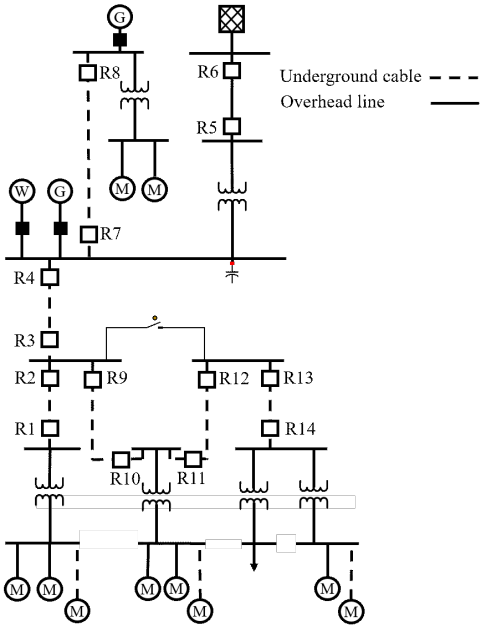



Figure 4. Modified 14-bus test feeder.

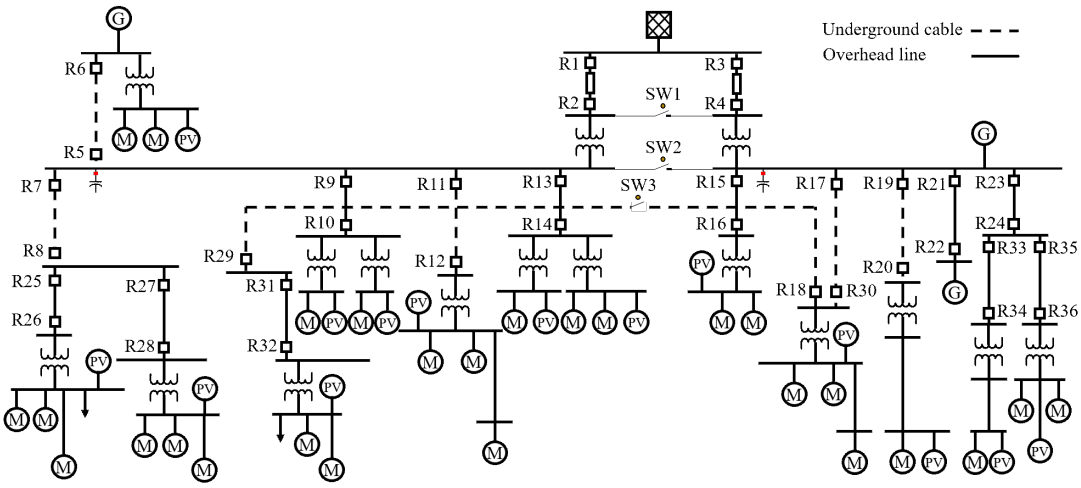



Figure 5. Modified 38-bus test feeder.

### *B. Convergence-guaranteed Learning*

*1) Short-circuit levels and some assumptions:* The results on 14-bus System focus on the algorithm itself instead of the overall setting, therefore, we demonstrate the short circuit current only for a selected branch in Table I. The system-level results can be found in the 38-bus system's results. The chosen branch has relays $R1$ to $R6$. Three-phase fault currents are demonstrated at the midpoint and near-end of the protected lines. In this section, we assume the DOCRs follow IEEE Inverse time curve (U2), with $A = 5.95$, $B = 0.180$, and $p = 2$. CTI is set to $0.3\,\text{sec}$.

Table I

SHORT CIRCUIT CURRENT OF THE SELECTED BRANCH IN THE 14-BUS SYSTEM.

| Prim. relay | Fault current ($A$) | | Bkp. relay | Fault current ($A$) | |
|---|---|---|---|---|---|
| | Midpoint | Near-end | | Midpoint | Near-end |
| 1 | 147.9 | 148.0 | - | - | - |
| 2 | 5334.3 | 5213.2 | 4 | 5334.3 | 5213.25 |
| 3 | 147.7 | 147.8 | 1 | 147.7 | 147.8 |
| 4 | 5522.3 | 5477.4 | 6 | 1085.6 | 1076.8 |
| 5 | 47.1 | 47.15 | 3 | 143.2 | 143.3 |
| 6 | 3886.2 | 3650.5 | - | - | - |

Practically, we deploy the relay element pickup ranges and accuracy of SEL-351 relay [3], as provided in Table II. To simplify the example, assume that the motors downstream of R1 have their own fault clearing time of $0.1\,\text{sec}$, thus, phase relay breaker near R2 has a maximum-operating time (MOT) for

far-bus fault of $0.1+0.3=0.4\,\text{sec}$. In the reverse direction, the MOT for breaker near R5 is set to $0.1\,\text{sec}$ as well.

Table II

RELAY ELEMENT PICKUP RANGES AND ACCURACIES.

| Setting | Min. | Max. | Step size | Note |
|---|---|---|---|---|
| TDS | 0.50 | 15.00 | 0.01 | IEEE curve |
| $I_{pu}$ | 0.25 A | 100.00 A | 0.01 A | 5 A nominal |

*2) Value table and its convergence:* Here, we zoom in to the highlighted row of Table I, focusing on the value table (also called Q-table) for R2 setting under various numbers of iterations and step sizes (resolution). The corresponding results are shown as heatmaps concerning its absolute values (all negative) in Fig. 6. The red blocks in Fig. 6(a) display to some extent the similarity with the value table structure in Fig. 3. The transaction in Fig. 6 illustrates two levels of change. One is that the increase of iteration numbers makes the values in Q-table smaller. This is intuitively correct since the rewards are all non-positive values; therefore, as the Q-table goes through more iterations, more negative values are added on the value table. The other transaction is from the impact of device accuracy. Commercial relays have a two-digit accuracy in TDS. Thus, the accuracy of $0.05$ and $0.01$ are tested. As Fig. 6(c) indicates, better selection of relay settings are available when the accuracy is higher.

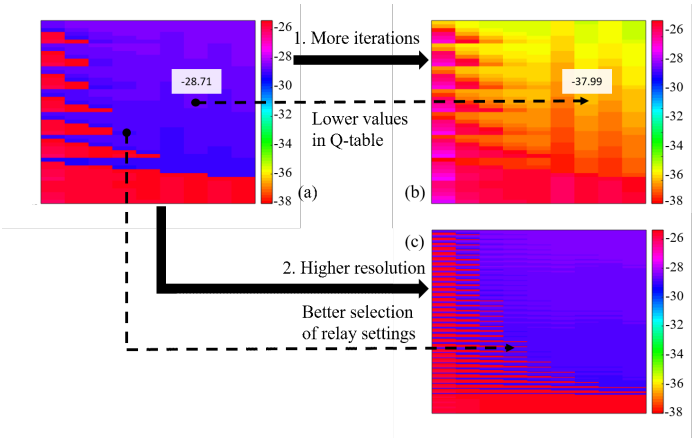



Figure 6. Value table for $R_2$ setting under four TDS step sizes. The step size of $I_{pu}$ is 20, maximum $I_{pu}$ is 100, minimum $I_{pu}$ is 0.25, step size of $I_{sc}$ is 600, maximum $I_{sc}$ is 5600, minimum $I_{sc}$ is 600. (a) The step size of TDS is 0.05, iteration 100. (b) The step size of TDS is 0.05, iteration 200. (c) Step size of TDS is 0.01, iteration 100.

The results in Fig. 7 provide evidence to the convergence of the proposed Q-learning setup. Two states are examined: state 1 follows the first condition in equation (3) that is slightly punished, and state 2 is

heavily punished with a coordination penalty ratio $b = 10$. It is observed from Fig. 7 that values for both states converge after $1,000$ iterations in this case.

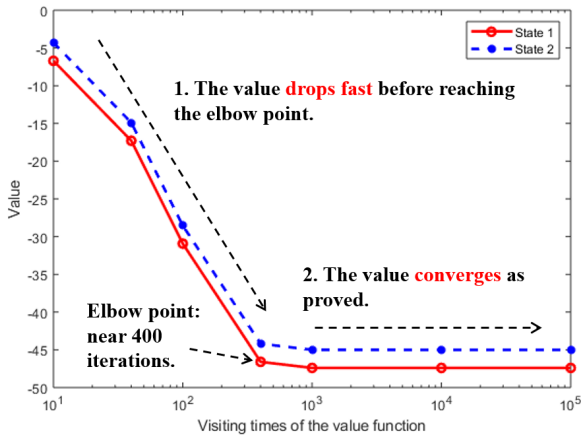



Figure 7. The change of iterative value with respect to the visiting times of the value function for two typical states in $R_2$ setting. State 1: TDS $= 1.50$, $I_{pu} = 100$ A, $I_{sc} = 2475$ A; State 2: TDS $= 2.50$, $I_{pu} = 1$ A, $I_{sc} = 2475$ A.

*3) Protective relay's action upon its observation:* Selected states and corresponding action values from Q-table in varying short-circuit environment are shown in Table III. We show $18$ cases of initial states and their post-division state. To better reveal the action taken, the column of intermediate action is listed, which is the subtraction of the aforementioned two states but does not mean we utilize them in the algorithm. Associated with the definition in Fig. 2, the first two elements are the direct actions $a_t$, and the third element is the void action $a_\emptyset$ (that is why they are always zero). In sum, the value table, convergence, and actions upon observation imply a convergence-guaranteed learning process.

### *C. High Adaptivity at System-level*

This larger system is employed to demonstrate the overall DOCR settings, the relay setting adjustment under the environment change, and the comparison with other methods. The overall protective relay settings based on our proposed Q-learning algorithm are shown in Table IV. The setting ranges and accuracy are in accordance with Table II. In this case study, the system assumes full load and DG generation. In the benchmark system, solar panels are installed on the roof of premises next to each load bus, e.g. the motor control room. The capacity of each solar system is assumed to be $80\%$ of the transformer capacity, taking the loss into consideration. Meanwhile, the short-circuit current contribution is set to $120\%$ of the rated current of PV inverters.

The proposed E-APS is capable of learning a large amount of scenarios due to the adoption of the reinforcement learning method. Here, we only show eight topological changes and DG on/off scenarios in a chronological order to emphasize the adaptability. The changes in topology are achieved through

Table III

SELECTED STATES AND CORRESPONDING ACTION VALUES FROM Q-TABLE IN VARYING SHORT-CIRCUIT ENVIRONMENT.

| Initial State | | | Post-decision State | | | Intermediate |
|---|---|---|---|---|---|---|
| TDS | $I_{pu}$ (A) | $I_{sc}$ (A) | TDS | $I_{pu}$ (A) | $I_{sc}$ (A) | Action |
| 1.90 | 1 | 600 | 2.14 | 21 | 600 | (0.24, 20, 0) |
| 2.15 | 100 | 1225 | 2.06 | 60 | 1225 | (-0.09, -40, 0) |
| 1.93 | 20 | 1850 | 2.09 | 80 | 1850 | (0.16, 60, 0) |
| 2.01 | 40 | 2475 | 2.18 | 60 | 2475 | (0.17, 20, 0) |
| 2.09 | 60 | 3100 | 2.21 | 401 | 3100 | (0.12, -20, 0) |
| 2.18 | 80 | 3725 | 2.22 | 40 | 3725 | (0.04, -40, 0) |
| 1.97 | 20 | 4350 | 2.21 | 60 | 4350 | (0.24, 40, 0) |
| 1.99 | 40 | 4975 | 2.22 | 40 | 4975 | (0.23, 0, 0) |
| 2.16 | 60 | 5600 | 2.21 | 100 | 5600 | (0.05, 40, 0) |

switches, some of which are normally open and some are normally closed. For example, when the transformer next to R2 is under maintenance, SW2 is closed to supply the left part of the network. As shown in Table V, a series of topology is presented by operating two switches. Meanwhile, four DG on/off scenarios, representing the percentages of connected PVs at $0, 50\%, 100\%$, and $0$ are presented. Since they are appearing in sequence, it emulates the PV short circuit scenarios when it is dark, half PVs are off, high irradiance, and then dark again. Fig. 8 shows the results highlighting the performance of E-APS at a particular relay (R18). Comparing with conventional DOCR setting method, the E-APS follows the change of the topology and DG status and therefore enhances the protective relay performance. The results also show a high adaptivity to achieve the optimal relay response time as the environment changes.

### *D. Broad Applicability Under Different Fault Types and Fault Resistance*

The proposed E-APS has been tested under different fault types and fault resistance. In this subsection, we demonstrate the results of R1-R6 using the 14-bus benchmark system as shown in Fig. 4. For the fault type study, we test the performance of the E-APS under three-phase-to-ground and single-line-to-ground faults. Not surprisingly, the single-line-to-ground fault currents have less magnitude comparing with those in Table I. Then, with the help of the E-APS, phase and ground protection settings are obtained in Table VI. It should be noted that the ground relay operates on $3I_0$ and is only subject to the unbalanced portion

Table IV

OVERALL PROTECTIVE RELAY SETTINGS OF THE 38-BUS SYSTEM IN AN EXAMPLE ENVIRONMENT.

| Relay | TDS | $I_{pu}$ (A) | Relay | TDS | $I_{pu}$ (A) |
|---|---|---|---|---|---|
| 1 | 4.17 | 0.25 | 19 | 0.83 | 0.25 |
| 2 | 0.83 | 0.25 | 20 | 2.50 | 0.25 |
| 3 | 5.84 | 0.25 | 21 | 0.83 | 0.25 |
| 4 | 0.83 | 0.25 | 22 | 2.50 | 0.25 |
| 5 | 0.83 | 0.25 | 23 | 2.50 | 0.25 |
| 6 | 2.50 | 0.25 | 24 | 2.50 | 0.25 |
| 7 | 2.50 | 0.25 | 25 | 0.83 | 0.25 |
| 8 | 2.50 | 0.25 | 26 | 4.01 | 9.30 |
| 9 | 0.83 | 0.25 | 27 | 0.83 | 0.25 |
| 10 | 2.50 | 0.25 | 28 | 4.04 | 9.07 |
| 11 | 0.83 | 0.25 | 29 | 3.90 | 8.25 |
| 12 | 2.50 | 0.25 | 30 | 2.50 | 0.25 |
| 13 | 0.83 | 0.25 | 31 | 0.83 | 0.25 |
| 14 | 2.50 | 0.25 | 32 | - | - |
| 15 | 0.83 | 0.25 | 33 | 0.83 | 0.25 |
| 16 | 2.50 | 0.25 | 34 | 4.04 | 15.99 |
| 17 | 4.17 | 0.25 | 35 | 0.83 | 0.25 |
| 18 | 2.50 | 0.25 | 36 | 3.90 | 9.26 |

Note: R32 is disconnected since line 31-32 is only $20\,\mathrm{ft}$.

of the feeder load. In the system under study, we set the phase loading unbalance to $20\%$. The results show that the proposed E-APS is applicable for both the phase and ground protection element setting.

To test the validity of the proposed method, we investigate the relay setting under various fault resistance. Determining the minimum fault current is a challenge, as fault currents can be below load levels. Certain assumptions are required because of the variability of faults on distribution systems. As suggested by [34], we can assume

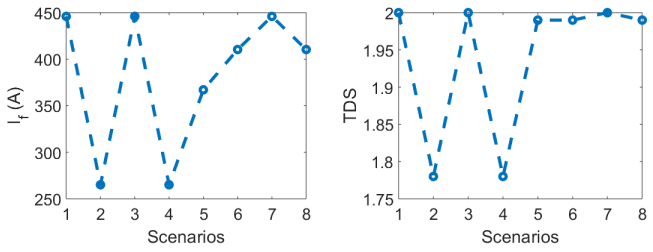
- some percentage of maximum fault current,
- some multiple of load current, and
- some fault impedance ($Z_f$) levels.

Table V

SELECTED TOPOLOGY UNDER STUDY.

| Topology | SW1 | SW2 | SW3 |
|---|---|---|---|
| T0 | Open | Close | Close |
| T1 | Open | Close | Open |
| T2 | Open | Open | Close |
| T3 | Open | Open | Open |

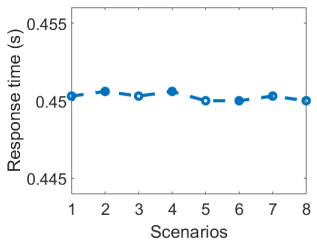
Note: To mitigate the circulating current between the two transformers near SW1 and SW2, SW1 is normally open.



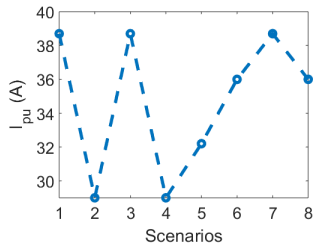

(a) Fault current.




(b) TDS.




(c) Response time.




(d) $I_{pu}$.

Figure 8. Environment-adaptive settings of R18 under eight scenarios. All PVs are assumed to be connected when topology changes, whereas the PV on/off status variation all occurs at T0 topology.

Hence, we have extensively tested the E-APS under various fault impedance. Table VII exhibits the E-APS performance of six phase elements with the three-phase-to-ground fault impedance of 0, 10, 20, and 30 $\Omega$, which is an empirical testing range guided by [34]. In this testing, we assume, upon the far-end faults, the operating times of the forward direction relay R2, R4, and R6 be at least 1.0 sec, 1.3 sec, 1.6 sec respectively, and the backward direction relay R5 be 1.0 sec (response time for R1 and R3 are not required in this scenario due to the system topology). The relay settings follow the ones in Table VI obtained through the E-APS. We list both the fault current seen by each relay and the operating time under the testing range. It is seen that the operating time remains a three-digit accuracy even though the

Table VI

PHASE AND GROUND OVERCURRENT PROTECTION SETTINGS OF 14-BUS SYSTEM IN AN EXAMPLE ENVIRONMENT.

| Phase protection | | | Ground protection | | |
|---|---|---|---|---|---|
| Relay | TDS | $I_{pu}$ (A) | Relay | TDS | $I_{pu}$ (A) |
| 1 | 4.99 | 19.02 | 1 | - | - |
| 2 | 5.56 | 0.50 | 2 | 0.84 | 0.25 |
| 3 | 2.03 | 46.59 | 3 | - | - |
| 4 | 7.23 | 0.25 | 4 | 2.50 | 0.25 |
| 5 | 11.01 | 24.70 | 5 | - | - |
| 6 | 5.01 | 45.94 | 6 | 4.01 | 38.24 |

Note: The ground elements of R1, R3, and R5 are not available due to the proximity to the delta side transformer connection.

fault impedance reaches up to $30\,\Omega$. Consequently, the relays respect the coordination rules along with other relays to achieve overall coordination using the E-APS.

Table VII

RELAY OPERATING TIME UNDER FAR-END FAULTS WITH DIFFERENT FAULT IMPEDANCE.

| Fault impedance ($\Omega$) | 0 | 10 | 20 | 30 |
|---|---|---|---|---|
| Fault current seen by R1 (A) | 6.7 | 3.8 | 2.2 | 1.5 |
| $T_{op,R1}$ (sec) | - | - | - | - |
| Fault current seen by R2 (A) | 44765.1 | 785.5 | 395.0 | 263.8 |
| $T_{op,R2}$ (sec) | 1.0008 | 1.0008 | 1.0009 | 1.0009 |
| Fault current seen by R3 (A) | 6.7 | 3.8 | 2.2 | 1.5 |
| $T_{op,R3}$ (sec) | - | - | - | - |
| Fault current seen by R4 (A) | 44765.1 | 785.5 | 395.0 | 263.8 |
| $T_{op,R4}$ (sec) | 1.3009 | 1.3009 | 1.3010 | 1.3011 |
| Fault current seen by R5 (A) | 1420.9 | 799.1 | 467.5 | 324.0 |
| $T_{op,R5}$ (sec) | 1.0007 | 1.0007 | 1.0007 | 1.0009 |
| Fault current seen by R6 (A) | 303.3 | 5.3 | 2.7 | 1.8 |
| $T_{op,R6}$ (sec) | 1.6009 | - | - | - |

### *E. Advantages Over Other Optimization-based Methods*

Further comparison is conducted to demonstrate the superiority of the proposed E-APS. As shown in Table VIII, the proposed method is compared with the conventional method and another optimization-based relay coordination method in five categories. The conventional method has several groups of relay settings that can be triggered by a certain logic, but it has poor adaptivity due to the device limitations. The method in [18] utilizes SCADA for system monitoring and relay setting update, while the proposed E-APS uses secure R-GOOSE mapping from IEC 61850 for its peer-to-peer communication. Additionally, while the E-APS has convergence guarantees and its results are repeatable, the method in [18] does not have an optimal solution guaranteed and cannot reproduce the same results. As for the computational power consumed, the E-APS is dependent on the number of states, which costs less compared with the method in [18] given the accuracy of modern relays.

Table VIII
COMPARISON WITH OTHER DOCR COORDINATION SCHEMES.

| Category | Conventional method | Method in [18] | E-APS |
|---|---|---|---|
| Adaptivity | Limited | High | High |
| Comm. protocol | Not necessary | IEC 61850 (SV and GOOSE) | IEC 61850 (R-GOOSE) |
| Optimality | Not guaranteed | Not guaranteed | Guaranteed |
| Reproduction of results | Yes | No | Yes |
| Computational power | Low | High | Medium |

## V. DISCUSSIONS

Through our numerical simulations, the E-APS has exhibited tremendous advantages. It monitors the system environment and intelligently adjusts its protection settings to achieve a high reward function in the long run. The E-APS is applicable in multiple systems, the core algorithm guarantees fast convergence during interactive learning, it is highly adaptive under topological changes and DG on/off status, and our test indicate that different fault types and fault resistance do not jeopardize the coordination. Lastly, compared to other methods, the E-APS has a superior communication protocol, optimality guarantees,

and result reproduction. However, the significant trade-off of achieving these advantages is the setup of an interactive environment between the system model and the protection agent. Off-the-shelf short-circuit analysis software is used to generate the outcome through the protection agent's effort on exploration and exploitation. The protection agent saves the learned policy for controlling the settings. What links the short-circuit analysis software and the agent is the software's API or scripting tool that is available in lots of software.

Moreover, the communication delay is insignificant for the proposed method. On one hand, the protection application of AC distribution networks is relatively less time critical as the fault current is not significant comparing with DC networks and transmission networks. On the other hand, the backup is taken good care of in the proposed method, which will be discussed later. Since the operating time of DOCRs varies from a fraction of a second to a few seconds and this paper focuses on the topological changes and the on/off status of DGs, the delay of around several tens of milliseconds is tolerable for this overcurrent protection scheme. Through literature review, the above findings are also evidenced by [27].

## VI. Conclusions

This paper realizes the environment-adaptive protection scheme with a reinforcement-learning based protective relay coordination method. Regarding the three key elements in reinforcement learning, void and direct actions are defined in the protection field, the states are systematically embedded in the value table, and the reward is carefully designed to reflect the relay response time. We not only introduce the concept of post-decision state into the value iteration method but also prove the convergence of Q-learning for relay setting coordination. The feasibility of the proposed E-APS is demonstrated through a $14$-bus and a $38$-bus system. Results serve as an evidence of the convergence of the adaptive protection scheme. Through comparison, the adaptivity of the E-APS presents a promising dynamic under topology and DG on/off status changes. In addition, the E-APS excels and achieves a good balance at adaptivity, communication protocol, optimality, reproduction of results, and computational power consumption.

## Appendix

### A. Proof of the Contraction Property

For any $x \in \mathcal{X}$, we have

$$
\begin{aligned}
&|BV_1(x) - BV_2(x)| \\
&= |\max_a \sum_y \mathcal{P}_A(f(x,a),y)[r + \gamma V_1(y)] \\
&- \max_{a'} \sum_y \mathcal{P}_A(f(x,a'),y)[r + \gamma V_2(y)]| \\
&= \gamma |\max_a \sum_y \mathcal{P}_A(f(x,a),y) V_1(y) \\
&- \max_{a'} \sum_y \mathcal{P}_A(f(x,a'),y) V_2(y)| \\
&\leq \gamma \max_a |\sum_y \mathcal{P}_A(f(x,a),y) V_1(y) - \sum_y \mathcal{P}_A(f(x,a),y) V_2(y)| \\
&= \gamma \max_a \sum_y \mathcal{P}_A(f(x,a),y) |V_1(y) - V_2(y)| \\
&\leq \gamma ||V_1 - V_2||_\infty \max_a \sum_y \mathcal{P}_A(f(x,a),y) = \gamma ||V_1 - V_2||_\infty
\end{aligned}
$$

## ACKNOWLEDGMENT

The authors would like to thank Salt River Project (SRP) for providing real load and solar generation data, and Dr. John Undrill for the inspiring discussions. This work was supported in part by the National Science Foundation of the United States under Grant 1810537, in part by the ARPA-E Project on "Sensor Enabled Modeling of Future Distribution Systems with Distributed Energy Resources", and in part by the Fonds de recherche du Québec — Nature et technologies project on "Protection des Réseaux de Distribution utilisant un Raisonnement Guidé par Données".

## REFERENCES

[1] Bloomberg New Energy Finance, "New energy outlook 2018," 2018, last accessed 20 February 2019. [Online]. Available: https://bnef.turtl.co/story/neo2018

[2] F. Katiraei, J. Holbach, T. Chang, W. Johnson, D. Wills, B. Young, L. Marti, A. Yan, P. Baroutis, G. Thompson *et al.*, "Investigation of solar pv inverters current contributions during faults on distribution and transmission systems interruption capacity," in *Western Protective Relay Conference*, 2012, pp. 70–78.

[3] SEL-351. (2016) Protection, automation, and control system. [Online]. Available: https://www.selinc.com/SEL-351/

[4] A. S. Meliopoulos, E. Polymeneas, Z. Tan, R. Huang, and D. Zhao, "Advanced distribution management system," *IEEE Transactions on Smart Grid*, vol. 4, no. 4, pp. 2109–2117, 2013.

[5] A. S. Meliopoulos, G. J. Cokkinides, Z. Tan, S. Choi, Y. Lee, and P. Myrda, "Setting-less protection: Feasibility study," in *2013 46th Hawaii International Conference on System Sciences*. IEEE, 2013, pp. 2345–2353.

[6] H. F. Albinali and A. S. Meliopoulos, "Resilient protection system through centralized substation protection," *IEEE Transactions on Power Delivery*, vol. 33, no. 3, pp. 1418–1427, 2018.

[7] S. Choi and A. S. Meliopoulos, "Effective real-time operation and protection scheme of microgrids using distributed dynamic state estimation," *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 504–514, 2016.

[8] A. M. Tsimtsios and V. C. Nikolaidis, "Towards plug-and-play protection for meshed distribution systems with DG," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 1980–1995, 2020.

[9] J. L. Blackburn and T. J. Domin, *Protective relaying: principles and applications*. CRC press, 2015.

[10] M. Ojaghi, Z. Sudi, and J. Faiz, "Implementation of full adaptive technique to optimal coordination of overcurrent relays," *IEEE Transactions on Power Delivery*, vol. 28, no. 1, pp. 235–244, Jan 2013.

[11] F. Coffele, C. Booth, and A. Dyśko, "An adaptive overcurrent protection scheme for distribution networks," *IEEE Transactions on Power Delivery*, vol. 30, no. 2, pp. 561–568, Apr 2015.

[12] Z. Liu, C. Su, H. K. Høidalen, and Z. Chen, "A multiagent system-based protection and control scheme for distribution system with distributed-generation integration," *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 536–545, Feb 2017.

[13] A. R. Haron, A. Mohamed, H. Shareef, and H. Zayandehroodi, "Analysis and solutions of overcurrent protection issues in a microgrid," in *IEEE Conference on Power and Energy (PECon)*, Dec 2012, pp. 644–649.

[14] H. H. Zeineldin, H. M. Sharaf, D. K. Ibrahim, and E. E. A. El-Zahab, "Optimal protection coordination for meshed distribution systems with dg using dual setting directional over-current relays," *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 115–123, Jan 2015.

[15] H. M. Sharaf, H. H. Zeineldin, and E. El-Saadany, "Protection coordination for microgrids with grid-connected and islanded capabilities using communication assisted dual setting directional overcurrent relays," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 143–151, Jan 2016.

[16] P. P. Bedekar and S. R. Bhide, "Optimum coordination of directional overcurrent relays using the hybrid ga-nlp approach," *IEEE Transactions on Power Delivery*, vol. 26, no. 1, pp. 109–119, Jan 2010.

[17] M. M. Mansour, S. F. Mekhamer, and N. El-Kharbawe, "A modified particle swarm optimizer for the coordination of directional overcurrent relays," *IEEE transactions on power delivery*, vol. 22, no. 3, pp. 1400–1410, Jul 2007.

[18] M. Y. Shih, A. Conde, Z. Leonowicz, and L. Martirano, "An adaptive overcurrent coordination scheme to improve relay sensitivity and overcome drawbacks due to distributed generation in smart grids," *IEEE Transactions on Industry Applications*, vol. 53, no. 6, pp. 5217–5228, Nov 2017.

[19] T. Amraee, "Coordination of directional overcurrent relays using seeker algorithm," *IEEE Transactions on Power Delivery*, vol. 27, no. 3, pp. 1415–1422, Jul 2012.

[20] D. Birla, R. P. Maheshwari, and H. O. Gupta, "A new nonlinear directional overcurrent relay coordination technique, and banes and boons of near-end faults based approach," *IEEE Transactions on Power Delivery*, vol. 21, no. 3, pp. 1176–1182, Jul 2006.

[21] IEC TR 61850-90-5, "Communication networks and systems for power utility automation - Part 90-5: Use of IEC 61850 to transmit synchrophasor information according to IEEE C37.118," International Electrotechnical Commission, Geneva, Switzerland, Standard, May 2012.

[22] J. R. Yellajosula, N. Sharma, M. Sundararaman, S. Paudyal, and B. A. Mork, "Hardware implementation of r-goose for wide-area protection and coordination," in *IEEE/PES Transmission and Distribution Conference and Exposition (T&D), Denver, CO*, Apr 2018.

[23] M. Kanabar, A. Cioraca, and A. Johnson, "Wide area protection & control using high-speed and secured routable goose mechanism," in *69th Annual Conference for Protective Relay Engineers (CPRE), College Station, TX*. IEEE, Apr 2016.

[24] A. Apostolov, "R-GOOSE: what it is and its application in distribution automation," *CIRED-Open Access Proceedings Journal*, vol. 2017, no. 1, pp. 1438–1441, Oct 2017.

[25] A. Apostolov, "To GOOSE or not to GOOSE? - that is the question," in *68th Annual Conference for Protective Relay Engineers, College Station, TX*, Mar 2015.

[26] M. Y. Shih, C. A. C. Salazar, and A. C. Enríquez, "Adaptive directional overcurrent relay coordination using ant colony optimisation," *IET Generation, Transmission & Distribution*, vol. 9, no. 14, pp. 2040–2049, 2015.

[27] P. P. Parikh, T. S. Sidhu, and A. Shami, "A comprehensive investigation of wireless LAN for IEC 61850–based smart distribution substation applications," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 3, pp. 1466–1476, 2012.

[28] T. Jaakkola, M. I. Jordan, and S. P. Singh, "Convergence of stochastic iterative dynamic programming algorithms," in *Advances in neural information processing systems*, 1994, pp. 703–710.

[29] S. Koenig and R. G. Simmons, "Complexity analysis of real-time reinforcement learning," in *AAAI*, 1993, pp. 99–107.

[30] A. G. Barto, S. J. Bradtke, and S. P. Singh, *Real-time learning and control using asynchronous dynamic programming*. University of Massachusetts at Amherst, Department of Computer and Information Science, 1991.

[31] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[32] C. Păduraru, D. Precup, J. Pineau, and G. Comănici, "An empirical analysis of off-policy learning in discrete mdps," in *European Workshop on Reinforcement Learning*, Jan 2013, pp. 89–102.

[33] K. Damron, "Distribution Device Coordination," 2016, last accessed 2020-05-04. [Online]. Available: https://na.eventscloud.com/file_uploads/22dbde83fcd108eb1d18a81c35c1a679_2016HORSAvistaDistributionPaper.pdf

[34] R. U. Service, "Design Guide for sectionalizing distribution lines, BULLETIN 1724E–102," 2012, last accessed 2020-04-27. [Online]. Available: https://www.rd.usda.gov/files/UEP_Bulletin_1724E-102.pdf