IEEE Open Access Journal of
Power and Energy

Received 16 November 2021; revised 8 February 2022; accepted 11 March 2022. Date of publication 24 March 2022;
date of current version 29 December 2022.

Digital Object Identifier 10.1109/OAJPE.2022.3161904

Deep Reinforcement Learning-Based Robust Protection in DER-Rich Distribution Grids

DONGQI WU[1] (Student Member, IEEE), DILEEP KALATHIL[1] (Senior Member, IEEE), MIROSLAV M. BEGOVIC[1] (Life Fellow, IEEE), KEVIN Q. DING[2] (Member, IEEE), AND LE XIE[1] (Fellow, IEEE)

[1]Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA
[2]CenterPoint Energy, Houston, TX 77002 USA

CORRESPONDING AUTHOR: L. XIE (le.xie@tamu.edu)

This work was supported in part by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) through the Solar Energy Technologies Office (SETO) under Grant DE-EE0009031, in part by the National Science Foundation under Grant OAC-1934675 and Grant ECCS-2035688, and in part by the Texas A&M Engineering Experiment Station Smart Grid Center.

ABSTRACT This paper introduces a new framework of deep reinforcement learning based protective relay design in power distribution systems with many distributed energy resources (DERs). With increasing penetration of power electronically-interfaced resources, conventional overcurrent relays' performance is rendered less effective due to the two-way uncertainties in power flow patterns. In this paper, a machine learning-based protective relay that is designed for adaptively deciding the threshold for relay action is proposed. The particular algorithm used is an Long Short-Term Memory (LSTM) enhanced deep neural network that is highly accurate, communication-free and easy to implement. The proposed relay design is tested in OpenDSS simulation on the IEEE 34-node test feeder and a collection of large synthetic feeders in Austin, Texas area. By designing adaptability upfront, the proposed relay is shown to substantially improve the performance of relay in terms of failure rate, robustness, and response speed, in particular in scenarios with high level of distributed energy resources.

INDEX TERMS Power distribution systems, protective relaying, reinforcement learning.

I. INTRODUCTION

THIS paper introduces a novel Deep Reinforcement Learning (Deep RL) based approach for robust protective relay control design in distribution grids. Recent developments in photovoltaic (PV) and power electronics technology have led to an increase of penetration of distributed energy resources (DER) in distribution grids. DERs, especially solar PVs, can provide a number of benefits to the power system operation efficiency such as peak load reduction and improved power quality [1]. However, DER and emerging grid edge-level devices are increasing the complexity of the interactions between end users and distribution grid operators substantially, such as low or non-existent system inertia, islanded operation and load-side voltage security. These additional complexities pose significant challenges for the operation and protection of the distribution grid.

Protective relays are the safeguards of distribution systems. The role of protective relays is to protect the grid from sustaining faults by disconnecting the smallest practically available faulty segment from the rest of the grid. During the operation, a relay monitors the power grid and looks for patterns that are associated with faults. Typical measurements include current (over-current and differential relay), voltage and current (distance relay), frequency or electromagnetic wave from transients (traveling-wave relay). In power distribution systems, time delayed, coordinated overcurrent relays are most commonly used since many other methods are impractical due to cost, infrastructure and grid topology limitations.

However, it is very difficult for overcurrent relays to accommodate the vastly different operational conditions in real distribution grids. For feeder recloser relays, the presence of DER within the feeder can reduce the fault current measured at the recloser and make faults harder to detect. The fault current contribution from DERs to the fault will also make the fault current observed at the fuse higher than the current at the recloser, making coordination based on inverse-time curves difficult [2], [3]. Moreover, even in current distribution grids, factors like fault impedance and load profile change are not taken into account in traditional overcurrent protection design, resulting in problems such as

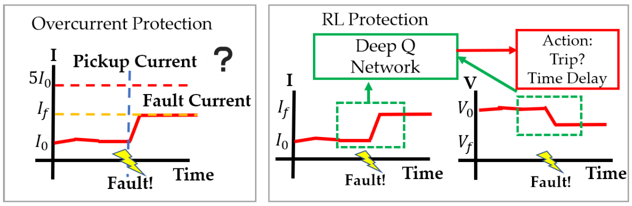


FIGURE 1. **Concept of Overcurrent and RL Protection.**

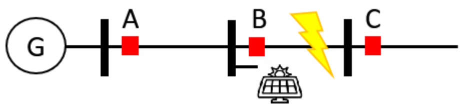


FIGURE 2. **Diagram of a Simple Distribution System.**

failing to detect faults near the end of a feeder, a.k.a. underreaching. Fig. 1 shows a conceptual comparison between threshold-based overcurrent protection and our proposed RL protection. Where overcurrent relays may be affected by low fault current magnitude, our method does not suffer the same limitation as it detects faults using the waveform patterns in measurements.

Conventional protective relays are also designed to function under two crucial assumptions: (i) power flow is unidirectional from the substation towards the end users, and (ii) the difference between operating conditions (currents and voltages) between normal and faulted conditions are measurable and significant. With the increasing popularity and penetration of DER and grid-edge devices, both assumptions will likely be rendered invalid [2]. For example, in the simple circuit shown in Fig. 2, there is a distributed generator feeding power into the grid at bus B. Under conditions where the net power absorption of the loads at bus B and C is low, or the output of the distributed generator at B is having a high peak, the power flow direction in the line between A and B will be from B to A, which violates assumption (i). For a fault to the right of bus B as indicated in the figure, the fault current contribution of the distributed generator could decrease the magnitude of fault current measured at bus A to the range of peak load current under normal conditions and potentially violating assumption (ii). In fact, reliable protection is becoming an Achilles' heel that limits the growth of DER integration for future grids.

Here we illustrate the challenges posed by DER on the protection system using a simple numerical example. Consider the same radial feeder in Fig. 2. This circuit has only one load at bus C, a distributed generator is placed at bus B. This generator is modeled as constant for the purpose of this example with current limit to roughly mimic an inverter based supply. The line parameters are adopted from the IEEE 4 bus feeder system. The rated current is calculated without the distributed generator and all power are supplied by the substation through the transmission grid. Under this simple illustrative setting, we vary the capacity of the distributed generator as a percentage of the total load, add a single-phase fault with various level of fault impedance and record the current at bus A. The ratio between fault current and load current is listed in Table 1.

TABLE 1. **Fault current level under various DER fraction of total load and fault impedance.**

DER % \ $Z_f(\Omega)$	0.01	0.1	1	5
0	3.841	3.676	2.197	1.283
25	3.875	3.653	1.995	1.074
50	3.913	3.629	1.794	0.877
75	3.956	3.603	1.596	0.697

It can be seen that the presence of distributed generator and fault impedance may greatly reduce the magnitude of fault current. Usually, for overcurrent relays, the fault current needs to be at least 2 to 3 times higher than the normal operation current under maximum load level to detect faults reliably. Under the impact of DER or fault impedance, or both, the fault current magnitude can be too low to detect for overcurrent relays. In contrast, for this simple example, we will demonstrate that our proposed RL relay is able to successfully detect faults under all scenarios above.

A. LITERATURE REVIEW

The improvement of protective relays in DER-rich systems has led to a large body of literature. Most of them focus on improving the performance of commonly used overcurrent relays by better fault detection [4] and coordination [5]. Neural networks have been used in setting the parameters of overcurrent relays [6]. Support Vector Machine (SVM) can be trained to distinguish the normal and fault conditions directly [7], [8]. A recent work [9] uses tabular Q-learning to find the optimal setting for overcurrent relays. Most proposed methods are still confined within the framework of inverse-time overcurrent protection, which is considered not enough for the future distribution grid with high DER and EV penetration [2].

Reinforcement Learning (RL) is a branch of machine learning that addresses the problem of learning optimal control policies for unknown dynamical systems. RL algorithms using deep neural networks [10], known as Deep RL algorithms, have made significant achievements in the past few years in areas like robotics, games, and autonomous driving [11]. RL has also been applied to various power system control problems including voltage regulation [12], frequency regulation [13], reactive power control [14], power quality control [15] and generator control [16]. Our previous paper [17] was the first work to use deep RL for power system protection. A comprehensive survey of RL applications in power system is detailed in a recent review paper [18].

B. MAIN CONTRIBUTIONS

In this paper, we present a novel deep RL based framework to design for robust protective relays in distribution grids with many DERs. Key contributions are suggested as follows:

- We formulate the design of protective relay in radial distributed systems as a nested RL problem.
- Based on the new formulation, we proposed a novel Long-Short-Term-Memory (LSTM)-enhanced RL algorithm that leads to much more reliable and accurate coordination of protective relays as compared to conventional inverse time over-current relays.
- We develop a fully automated software interface that complies with OpenAI Gym that is readily available to integrate state-of-the-art machine learning packages and commercial grade power system simulators.

This work is substantial expansion of the preliminary work reported in [17]. Major improvements includes detailed formulation to three-phase unbalanced scenarios, an LSTM enhanced network model, a diverse set of test scenarios involving large realistic systems. This paper is organized as follows: Section II formulates the protection problem using the RL framework and introduces our nested reinforcement learning algorithm. Section III presents the simulation environment and test-beds used in training and evaluation, Section IV analyzes and discusses simulation results. Section V summarizes and concludes the paper.

II. PROBLEM FORMULATION

A. RELAY OPERATION

Here we briefly present typical operation of protective devices using a simple distribution line in Fig. 2. There are 3 protective devices marked as red boxes along the line. The purpose of these devices is to protect the upstream from short-circuit fault current by breaking the electric connection using a circuit breaker. The upstream protection device closest to the fault (i.e. red box B for the fault shown in the diagram) should trip to disrupt as fewer load as possible. The protective devices are designed to coordinate with each other to provide backup protection to improve reliability. Typically, there are two general coordination setups for different types of distribution systems. In long mid-voltage sub-transmission lines where multiple relays are installed along the line, the relays are programmed with different inverse-time curves such that they trip with different time delays. In the above example, the trip delay for relay A is longer than that of B by a fraction of a second, to allow the closer relay B to trip first; Relay A will only trip if B fails to work properly. In end-level low voltage distribution feeders it is more common to employ a recloser-fuse coordination in which only one relay is placed near the source (A) and the rest of the systems are protected by fuses (B and C). It is desired that after faults the recloser would perform reclosing operations before the fuses melts for transient faults, as the melting of fuses is irreversible and introduce addition cost. This coordination is usually implemented using slow-fast curves such that the recloser attempts to clear transient faults by quickly opening and reclosing, and if the fault is persistent, the fuse will melt to clear the fault shortly after. If the fuse fails to melt, the recloser will be locked open as backup protection.

B. REINFORCEMENT LEARNING AND MODELING OF RL RELAY

In RL formulations, a control problem is modelled as an active interaction between the controller, a.k.a. *Agent*, and the system, or environment, to be controlled. The system is represented by a Markov Chain whose state evolves based on a deterministic or stochastic transition kernel as well as the actions of the agent. The agent observes the state of the system and give control actions based on its *policy*. Each action is assigned a *reward* that is based on the effect of the action and the resulting state transition. In the process of solving an RL problem, the agent learns a control policy that gives the most optimal action corresponding to each observed system state in order to maximize total expected reward. Unlike traditional control problems in which the controller is derived from analytical development of an explicit model of the plant, an RL agent learns its optimal policy through extensive observation under perturbations of the system state. The RL agent typically assume no prior knowledge about the system model at the beginning of learning, it then gather experience about the system state transition and reward by attempting different actions under different states. After enough experience is collected, the agent will be able to choose the actions that results in the highest long-term reward based on the observations it receives.

We propose a RL based relay control strategy that is adaptive and robust even in DER-rich distribution grids. The RL relay can take the same or more measurements available to traditional overcurrent relays and output a tripping signal with a time delay for coordination without communication when it detects a fault. Unlike the threshold-based fault detection logic that are designed from a finite set of scenarios and strong assumptions, the RL relay can learn a policy that is based on not only the instantaneous post-fault current, but also pre-fault condition and system dynamics from the transient response. A flow diagram for comparison of the concept of standard overcurrent relay and our RL relay design is shown in Fig. 3. During the training of the RL relay policy, the agent explores a large number of current measurement around the time of fault events and learn the pattern associated with those events as well as the correct response. To facilitate the training process, a synthetic model is needed to produce the large amount of training data. This is achieved by building a simulation environment that can generate random fault scenarios and adjust according to the agent's actions. The minutiae of implementation are discussed in section 3.

This formulation using reinforcement learning, compared to other machine learning based methods, has several advantages that are especially appropriate for the decision making of protective relays. First, most methods such as support vector machine or artificial neural network take a supervised learning approach, which attempts to develop a best classifier to distinguish normal and fault conditions from the training data-set. These methods require all training data to be properly labeled in advance which, in this case,

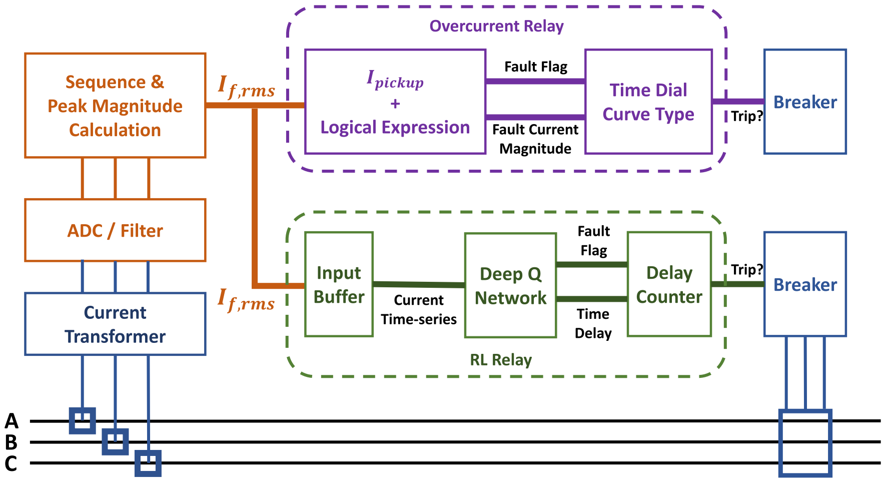


FIGURE 3. **Conceptual Comparison Diagram between Overcurrent and RL relay.**

means the optimal relay action and associated time delay for each series of current measurement in the training data. Although supervised learning methods work reasonably well in identifying faults, it state-less structure is inherently not compatible with consecutive and inter-dependent decisions in online control problems. In contrast, in RL explicit labeling of the training data is not required. The agent only need to be told that if the action it performs is desirable, which can be easily determined based on the status of the circuit. In short, supervised learning is an instructive process where the teacher need give step-by-step instructions to tell the agent exactly what to do during training, while RL is an evaluative process where the trainer only need to provide an evaluation of the actions taken by the agent based on the final outcome. The latter is more suitable in a dynamic control problem such as in the electric grid where typically a well-formulated objective need to be achieved. Second, it can be difficult to incorporate the underlying model of the system in other data-driven techniques. Consecutive measurement taken at the current/potential transformer by relays are determined by the network model, which is usually very complex or computationally intensive to obtain and utilize. However, an underlying system model is included in the formulation of RL, and during the learning process it learns the patterns in state transition and evolution in order to make a series of consecutive actions to achieve a desirable final objective.

C. MATHEMATICAL FORMULATION OF MARKOV DECISION PROCESSES AND REINFORCEMENT LEARNING

Next, we will proceed to give a brief review of the basic concepts of Markov Decision Process (MDP) and RL and then present a mathematical formulation of the protective relay problem. This formalism will be expanded later to formulate the optimal control for relay protection problem under the framework of multi-agent reinforcement learning. A concise but more comprehensive introduction of MDP, Dynamic Programming (DP) and RL could be found at [19].

Markov Decision Processes (MDP) is a mathematical framework for stochastic control problems. This framework models a control problem as a sequential decision making problem where the environment varies partly random and partly based on the control actions. An MDP is modeled as a *tuple* with 5 elements: $(\mathcal{S}, \mathcal{A}, R, P, \gamma)$ in which $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space. $P = (P(\cdot|s, a), (s, a) \in \mathcal{S} \times \mathcal{A})$ is the transition probabilities that corresponds to probability of transitioning to state s' from state s as a result of action a. $R : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is the reward function and $\gamma \in [0, 1)$ the discount factor. Under the RL framework, for a protective relay, its state space S will include all information available for observation (e.g. voltage, current or frequency); its action space A will include all possible breaker operation; P will be determined by the model of the distribution grid where the relay is deployed; reward R and discount factor γ will be chosen before training to promote desirable operations.

Under each state, the action of an agent is given by its policy $\pi : \mathcal{S} \rightarrow \mathcal{A}$. A policy is usually evaluated using the *value function*, V_π:

$$V_\pi(s) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t R_t | s_0 = s],$$

where $R_t = R(s_t, \pi(s_t))$ and $s_{t+1} \sim P(s_t, a_t)$. The *optimal value function*, V^* is the value function of the optimal policy that gives the highest value function: $V^*(s) = \max_\pi V_\pi(s)$. The optimal policy π^* can also be obtained from V^* using the Bellman equation:

$$\pi^*(s) = \arg\max_{a \in \mathcal{A}} (R(s, a) + \gamma \sum_{s' \in \mathcal{S}} P(s'|s, a) V^*(s')).$$

The Q-value function of a policy π, Q_π, is defined as $Q_\pi(s, a) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t R_t | s_0 = s, a_0 = a]$. The Optimal Q-value function Q^* is also defined similarly, $Q^*(s, a) = \max_\pi Q_\pi(s, a)$. The optimal policy could be directly obtained using the optimal Q-value function as $\pi^*(s) = \arg\max_{a \in \mathcal{A}} Q^*(s, a)$ Specifically, the optimal Q-value for tripping the breaker will be the highest when a fault is present on the feeder, which improves the reliability of relays; the optimal Q-value for staying closed will be the highest when the system is under normal conditions, which enhances the dependability of relays.

For MDP formulations, the optimal value/Q-value function or the optimal control policy can be computed using dynamic programming methods [20]. These explicit methods requires the full transition probability matrix P for all possible state variable combinations. However, in realistic problems the exact system model is usually difficult to obtain. Specifically, in the protective relay problem, the transition probability represents all possible stochastic variations in the phasor voltage and current in the network caused by uncertainties in load profile and DER generation. Fault parameters such as fault impedance and location are also rendering difficult accurate determination of that probability because some parameters of the model can vary widely from one fault to another. Learning based methods are generally more appropriate for such problems with high level of uncertainties.

Reinforcement learning is a method for learning the optimal policy for an MDP when a explicit model is not available. In RL, the optimal policy is learned through sequential observations and interactions with the system. In *Q-learning*, which is the most commonly used algorithm for RL, the optimal Q-value function is learned from a sequence of interactions (s_t, a_t, R_t, s_{t+1}). Specifically, the Q-value function Q_t is updated at each time step t as:

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha_t [R_t + \gamma \max_{b \in A} Q_t(s_{t+1}, b) - Q_t(s_t, a_t)] \quad (1)$$

where α_t is the step size, or learning rate, that determines how fast newly collected information gets incorporated to the existing model. With an appropriately chosen α, Q_t will converge to the optimal Q-function Q^* after each state-action pair has been visited sufficiently often [20].

The standard Q-learning algorithm as described cannot be directly used in problems with continuous state/action space. For continuous problems, a deep neural network is usually used as a replacement for an explicit Q-function: $Q(s, a) \approx Q_n(s, a)$ and n represents parameters of the neural network. The neural network for Q-learning is usually called Deep-Q-Networks (DQN).The ability of neural networks to *approximate any function* using only input-output samples has enabled tremendous success in many reinforcement learning problems in different fields.

For each state-action pair (s_t, a_t, R_t, s_{t+1}), the parameters of the DQN can be updated using stochastic gradient descent:

$$n = n + \alpha \nabla Q_n(s_t, a_t) (R_t + \gamma \max_{b \in A} Q_n(s_{t+1}, b) - Q_n(s_t, a_t)) \quad (2)$$

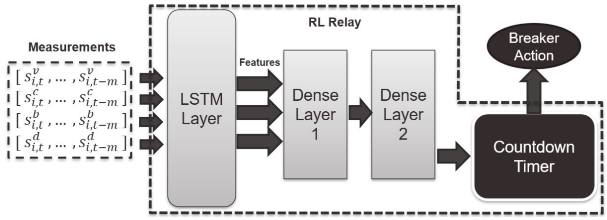


FIGURE 4. Model Structure of an RL Relay.

Q-learning with neural network can be improved by implementing various upgrade techniques. *Experience replay* is used as a buffer to store a batch of observations and shuffle them before each gradient upgrade. This can help to avoid the bias introduced by the temporal correlation among observations obtained in a sequence. *Target network* is a separate neural network model only used to temporarily fix the gradient descent target for several steps to avoid potential instability caused by chasing a moving target.

An Long-Short-Term-Memory(LSTM) layer [21] is used before the fully-connected layers to extract features from time series inputs. Using LSTM with deep reinforcement learning [22] has received increasing attention in recent years in time-correlated control problems. LSTM has a unique advantage over other non-recurrent neural network models, that is, the ability to remember what has happened in the past. Each LSTM cell has a internal state that can be either kept/changed/forgot for every observation it receives. This feature is particularly useful for assessing the current state of power systems, as it is able to adapt to the change of states incurred by other disturbances that does not need protection to operate (e.g. daily load curve, renewable generation profile, etc.). Our algorithm is based upon the combination of deep neural networks, experience replay, target network and LSTM feature extraction as illustrated by the flowchart in Fig. 4.

D. PROTECTIVE RELAY CONTROL AS AN RL PROBLEM

We formulate the distribution system transient process as an MDP environment and model the relays as RL agents. For consistency with the current protection infrastructure, each relay is set to only observe its local current measurements ($s^c_{i,t}$), although if additional information(voltage, frequency, etc.) is added in to the state space the RL relay would easily accommodate them without changing the formulation and potentially achieve even better performance. Each relay knows the status of the local current breaker circuits, i.e., if it is open or closed ($s^b_{i,t}$). Each relay has a local counter that ensures the necessary time delay in its operation as a backup relay ($s^d_{i,t}$). These variables constitute the state $s_{i,t} = (s^c_{i,t}, s^b_{i,t}, s^d_{i,t})$ of each relay i at time t. Table 2 summarizes this state space representation. Each state also uses the past m measurements to form a timeseries of measurement with length $m + 1$. An appropriate combination of sampling rate and length of the timeseries allow one to deal with some

TABLE 2. Relay State Space.

State	Description
$s^c_{i,t}$	Local current measurements of past m timesteps
$s^b_{i,t}$	Status of breaker (open (0) or closed (1))
$s^d_{i,t}$	Value of the countdown timer

TABLE 3. Relay Action Space.

Action	Description
a_{set}	Set the counter value to the desired number of delay steps
a_d	Decrease the value the counter by one
a_{reset}	Stop and reset the counter

classes of transients that cannot be identified from phasor measurements (such as inverter controls and limiters) in order to determine post-transient state.

Relay should operate after faults occur. However, since each relay is able to observe only its local state and no communication is assumed between the relays, some implicit coordination between relays is necessary. In traditional overcurrent protection scheme, the coordination is achieved using inverse-time curves that add a time delay between the detection of fault and actual breaker operation, based on the variation among fault current magnitudes at different locations of the circuit. However, fault current magnitudes can be unpredictable across different scenarios, especially with DER and smart edge-devices. We propose another approach (that is also amenable to RL) as follows. Instead of tripping the breaker instantaneously, it controls a countdown timer to indirectly operate the breaker. If a fault is detected, the relay can set the counter to a value such that the breaker trip after a certain time delay. The counter could be cancelled prematurely if the fault is cleared by another protective device. The action of each relay i at time t, $a_{i,t}$ is summarized in table 3.

The reward given to each relay is a measure of success for its most recent action. A positive reward is given to an RL relay if: i) it remains closed during normal conditions, ii) it trips the breaker after a fault in the downstream circuit where it is the closest protection device, or when other closer protections fail to operate. A negative reward is given if: i) tripping the breaker when there is no fault, or the fault is outside of its assigned region; 2) fail to trip the breaker when a fault is present in its assigned region. The magnitude of the rewards are designed to implicitly signify relative importance of false positives (lack of dependability) and false negatives (lack of reliability). The reward function for each relay is shown in Table 4.

The transition probability in a distribution feeder with multiple RL relays relates the change in power flow states to the measurement and operation of RL relays. Formally, let the global state at time t, $\bar{s}_t = (s_{1,t}, s_{2,t}, \ldots, s_{n,t})$, denote all nodal voltage and branch current in the system; let the combined action at time t, $\bar{a}_t = (a_{1,t}, a_{2,t}, \ldots, a_{n,t})$, denote the action of every RL relay in the system. Then, the state

TABLE 4. Reward for Different Operations.

Reward	Condition
Large Positive	Tripping when a fault is present in its assigned protection region
Large Negative	Tripping when there is no fault or the fault is outside its assigned region
Small Positive	Stay closed when there is no fault or the fault is outside its assigned region
Small Negative	Stay closed when a fault is present in its assigned protection region

of the system $\bar{s}_t$ evolves stochastically based on $\bar{a}_t$ plus the variation in load profile, DER output and circuit connectivity. Note that the global state evolution cannot be described by local transition probabilities of individual relays because the action of any relay can affect the states of other relays. The global system dynamics is represented by the transition probability $\bar{P}(\bar{s}_{t+1}|\bar{s}_t, \bar{a}_t)$.

The goal in the multi-agent RL formulation is to achieve a global optimum which maximizes the expected sum of reward received by all relays, using only local control laws π_i on local observations $s_{i,t}$: $\max_{(\pi_i)^n_{i=1}} \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t \bar{R}_t]$, $a_{i,t} = \pi_i(s_{i,t})$. Local policies π_i needs to be computed individually as a centralized policy would not be possible due to lack of communication.

E. ALGORITHM AND EXPANSION FOR MULTI-AGENT PROBLEMS

Some distribution systems have multiple active protection devices coordinating with each other. However, obtaining the policies for a network of distributed RL relays operating in the same system could be difficult because, 1) Normal RL methods require the environment to appear stationary to the agent; 2) The whole system state in a power grid is not observable using measurements collected from only one location. Multi-Agent-RL(MARL) [23] problems are often untrackable and the performance of available algorithm is generally not reliable.

We proposed a *Nested Reinforcement Learning* algorithm in Algorithm 1 that cleverly takes advantage of the radial structure of distribution systems to simplify the otherwise difficult MARL problem. In radial distribution systems, the dependency between the operation of coordinating relays is uni-directional, i.e., only upstream relays need to provide backup for a downstream relay but not vice-versa. Also, the last relay at the load side does not need to coordinate with others. In our nested RL algorithm, we start the RL training from the the most remote relay from the distribution transformer whose ideal operation is not affected by the operation of other relays, thus can be trained using a single-agent algorithm. Then, we can fix the trained policy for this last relay and train the relays at one-level closer to the substation that need to provide backup for the last relay. Since the policy of the furthest relay is fixed, it appears like a part of the stationary environment to its upstream neighbors which can learn to accommodate its operation. This process can be repeated for all the relays upstream to the substation. This method is analogous to how the coordination of time-delayed overcurrent relays is performed. The order of training can be determined by network tracing using a *post-order depth-first* tree traversal with the substation being the root. This nested training approach which exploits the nested structure of the underlying physical system allows us to overcome the non-stationarity in generic multi-agent RL settings. This approach is also used to help the RL recloser relay learn to coordinate with fuses by learning their operation patterns during training to achieve a fuse-saving scheme.

Algorithm 1 Nested Reinforcement Learning Algorithm

Initialize DQN of each relay i with random weights
Sort all relays based on system topology
for relay $i = 1$ to n **do**
  for episode $k = 1$ to K **do**
    Initialize simulation with random system parameters
    for time step $t = 1$ to T **do**
      Observe the state $s_{i,t}$ for all relays
      for relay $j = 1$ to i (Trained Relays) **do**
        Select action using the trained policy as:
        $a_{j,t} = \arg\max_a Q_{n_j^*}(s_{j,t}, a)$
      end for
      for relay $j = i + 1$ to n **do**
        Select do nothing action, $a_{j,t} = 0$
      end for
      With probability ϵ select a random action $a_{i,t}$, otherwise select the action with the highest Q value:
      $a_{i,t} = \arg\max_a Q_{n_i}(s_{i,t}, a)$
      Observe reward $R_{i,t}$ and next state $s_{i,t+1}$
      Store $(s_{i,t}, a_{i,t}, R_{i,t}, s_{i,t+1})$ in the replay buffer of relay i
      Sample a batch of past transitions from replay buffer and update the DQN parameter w_i
    end for
  end for
end for

III. EXPERIMENT ENVIRONMENT AND TEST CASES

A. SIMULATION ENVIRONMENT

The simulation environment is built by packing the OpenDSS APIs in a Python class inherited from the OpenAI Gym [24] to improve accessibility. We note that this setting can potentially be used in a number of other research problems addressing the distribution systems operation using machine learning. The RL algorithm is programmed in Python using open-source machine learning packages Tensorflow [25]. The hyper-parameters of the DQN for each relay are selected through random search are are listed in Table 5 to serve as a starting point for potential replications of the works.

B. TEST SYSTEM MODELING

We first use a common benchmark, the IEEE 34-bus test feeder (Fig. 5), to test the performance of RL based recloser relay control. The test cases are replicated in OpenDSS using the same parameters provided in IEEE publications [26]. Overall, OpenDSS power flow result and IEEE results agree closely, while the difference is mainly caused by aggregating distributed loads in a dummy bus at the midpoint of each branch. The percentage difference of node voltages between the OpenDSS simulations and IEEE published values are listed in Table 6. We also test our RL relay algorithm in a collection of large realistic synthetic distribution feeders in Syn-Austin-TDgrid-v03 [27]. This dataset contains 448 feeders around 140 substations in Austin, TX, most are low-voltage end level distribution circuits with detailed fuse configurations. The RL recloser relay is trained to coordinate with the fuses specified in the dataset for each feeder. Since the test cases do not specify the short-circuit current capacity of the source, the baseline values are selected accordingly from the IEC 60076-5 standard based on the voltage level of each circuit.

Modifications to the original cases are done when initializing each *episode* to simulate the real fluctuations of distribution grids. An episode is defined as a short simulation segment that contains a fault. A scenario is generated for each episode using a Monte-Carlo approach with a random combination of load and DER generation profile, fault parameter and fault location. The load and DER generation capacities are sampled from the COVID-EMDA+ dataset [28], which has the real hourly renewable generation and load data for cities within each RTO region. In the beginning of each episode, a random hour is chosen from the year 2019, and the recorded load profile and PV

TABLE 5. DQN Hyper-parameters.

Hyper-parameter	Value
LSTM Cell Number	70
Hidden Layers	256/128
Activation	ReLU/ReLU/Linear
Target Network Update Rate	0.005
Optimizer and Learning Rate	Adam, 0.0001

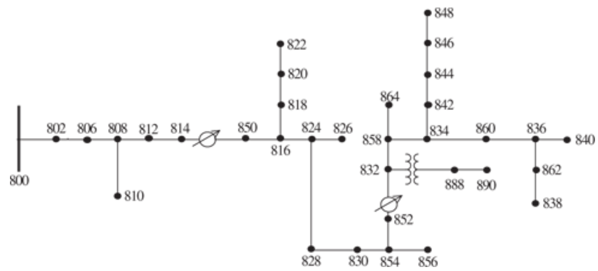


FIGURE 5. IEEE 34 node test feeder.

TABLE 6. Difference Between OpenDSS and IEEE Solution.

% Error	V_a	V_b	V_c
Average	0.179	0.240	0.023
Maximum	0.637	0.554	0.066

TABLE 7. Performance of Overcurrent Relays in Base Case.

Type	Occurrences	Probability
IEEE34		
False Alarm	0 / 5000	0 %
Fail to Detect	45 / 5000	0.9 %
Synthetic Austin		
False Alarm	0 / 5000	0 %
Fail to Detect	168 / 5000	3.36 %
Failed Coordination	494 / 5000	9.88 %

capacity for Houston, Texas corresponding to that time is used to scale the load and PV generators. The locations are randomly scattered throughout all single-phase loads. The randomization of DER placement is only meant to provide singular experimental scenarios, although we are aware of the fact that the placement will have an impact on relay performance and would require more thorough analyses. For larger systems, techniques in [29] could be used to reduce the amount of computation power required.

In the middle of an episode, a random fault is added to the system. The fault will occur in a random line and phase(s), have a random impedance from 0.001 ohm to 20 ohm. All types of faults (SLG, LL, LLG, 3-phase) are possible. To match realistic scenarios in distribution lines, single phase faults have the highest chance to be selected and 3-phase faults have the lowest chance. The performance of the RL relays are evaluated by running a large number of random episodes. In the following demonstration, 5000 independent episodes was used for evaluation under each type of scenario.

C. OVERCURRENT PROTECTION

To establish a baseline for comparison, a simple overcurrent recloser is placed at the substation and is configured to respond to faults in the distribution feeder. The settings of the overcurrent recloser relay is assumed to be twice the nominal current under the base case with a time dial of 0.1, in which the load capacities are the same as original numbers in the test cases and the substation transformer is the only power supply for the feeder. Since load profiles are normalized between 0 and 1, the original load capacities will be the maximum load in all simulated scenarios.

The fault detection of the overcurrent relays are tested under the basic IEEE 34 node feeder and one synthetic Austin feeder without considering any DER or load variation. The results in Table 7 shows the overcurrent relays are quite reliable under the static environment for a simple circuit. More specifically, the few times the overcurrent relays fail to detect faults in the IEEE 34 bus case are for single-phase faults in the 4.16 kV buses (888 and 890) with a relatively high fault impedance. However, for a larger system (the one used has 379 buses) with many branches, the performance of overcurrent protection becomes less desirable.

IV. SIMULATION RESULTS

A. PERFORMANCE METRICS

In this section we present and discuss the performance of our Nested RL algorithm for protective relays. We compare the performance with conventional overcurrent relay protection strategy. The performance is evaluated in three aspects:

Failure Rate: A relay failure happens when a relay fails to operate as it is expected to do. For each episode, we determine the optimal relay action from the type, time, and location of the fault, and compare it to the action taken by the RL based relay. We evaluate the percentage of the operation failures of the relays from 3 different aspects: when there is (i) no fault in the system; (ii) detection of faults; (iii) coordination with other protection equipment.

Robustness: The load profiles in power distribution systems is a combined result from factors including renewable generation, load ramping, weather and social events. Moreover, both the total load capacity and renewable penetration are expected to grow consistently each year. Increase in the load capacity can cause a higher peak load and high renewable penetration can increase the variance of the load profile. The rating and capacity of transmission and substation equipment, which is reflected in simulation as the short-circuit capacity (SCMVA), also changes alone the upgrade and reconfiguration of transmission system and substation equipment. It would be desirable if the protection system is robust against such changes to avoid the additional cost introduced by re-analyzing and re-programming the relays after deployment. We evaluate the performance of RL relays when the operating condition exceeds the nominal range.

Response Time: The response time of RL relay is defined as the time difference between the inception of the fault and the relay action. Response time is extremely critical in preventing hazards from cascading failures as well as saving fuses from melting unnecessarily. We compare the response time of the RL based relays with the conventional overcurrent relays.

TABLE 8. Failure Rate of Relays Under 30% DER in IEEE 34 Bus Feeder.

Scenario	False Operation	Occurrences	Probability
RL Based Relay			
No Fault	Trip	0 / 5000	0.00 %
Faulted	No Response	24 / 5000	0.48 %
Overcurrent Relay			
Scenario	False Operation	Occurrences	Probability
No Fault	Trip	0 / 5000	0.00 %
Faulted	No Response	826 / 5000	16.52 %

B. PERFORMANCE: BENCHMARK CIRCUIT - IEEE 34

We first test the fault detection performance of our RL algorithm for a single recloser control in a common benchmark system. The IEEE 34 bus feeder as shown in Fig. 5 is used in this experiment. The simulations are conducted separately with overcurrent protection and RL protection programmed in the same simulation setting. The simulation is run for 5000 randomly generated episodes and the operation of RL relay and overcurrent relay is logged and compared.

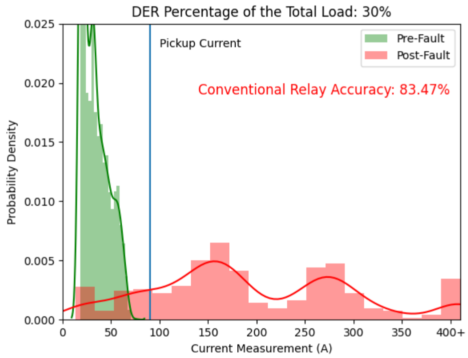


FIGURE 6. Pre-fault and Faulted Current Distribution at Bus 800, Estimated from Samples Recorded in 5000 Scenarios.

TABLE 9. Robustness Against Peak Load and DER Increase in IEEE 34 Bus Feeder.

Peak Load Increase	+10 %	+20 %	+30 %
RL Failure Rate	0.22 %	0.16 %	0.04 %
Overcurrent Failure Rate	3.28 %	2.76 %	2.34 %
Peak DER Increase	+10 %	+20 %	+30 %
RL Failure Rate	0.52 %	0.56 %	0.62 %
Overcurrent Failure Rate	20.96 %	22.36 %	23.8 %
SCMVA Variation	60 %	80 %	120%
RL Failure Rate	0.52 %	0.48 %	0.50 %
Overcurrent Failure Rate	17.12 %	16.80 %	16.02 %

Table 8 summarizes the **failure rate** performance of both the RL relay and overcurrent relay in 34 bus test feeder. The RL based relays are extremely accurate even under very high DER penetration levels. The fault current contribution from DER and fault impedance can, under many cases, reduce the magnitude of fault current measured at the substation (bus 800) considerably. As shown in Fig. 6, the fault current magnitude can be very close tho the normal load current range for faults near the end of the feeder, high-impedance faults or faults in the two 4.16kV buses. Under these scenarios, a fixed pickup current can never completely separate the normal and fault condition because their distributions are overlapping.

To quantify the **robustness** of RL based algorithm against peak load variations, the total load capacity the system is increased to up to 30% more than the peak capacity used to generate the training data. In creating the validation data for robustness assessment, we focus on the robustness only when the system load is around the peak. For evaluating the robustness at 10% higher load, the data is only selected when the system load is between 100% and 110% of the original capacity. Note that the model and policy of the RL relay remain unchanged, which means the data samples at the higher load are not used in training. The performance of RL and overcurrent relay under higher peak is shown in Table 9.

Similarly, we evaluate the robustness against potential increases in DER penetration and variation in SCMVA rating of the source. As the capacity of DERs in distribution systems is expected to increase over time, it is desirable that protection devices can reliably function without the need to re-configure their settings. In this experiment, the RL relays are trained using data created assuming an up to 30% DER penetration as described in Sec. IV, B. The obtained policy is tested under scenarios where the DER penetration is increased above 30%. The results are shown in the bottom half of Table 9. In all episodes where RL relay failed, the fault is located in the two 4.16 kV buses 888 and 890 with a relatively high fault impedance. In these cases, the rising edge in the fault current measured by the RL relay is not distinct enough to be detected. This location dependency may explain why the RL failure rate is not sensitive to change in DER and SCMVA level.

TABLE 10. RL Response Speed in IEEE 34 Bus Feeder.

Delay	1 Step	2 Step	3 Step	4 Steps
Occurrences	0 / 5000	4981 / 5000	17 / 5000	2 / 5000

We also measure the **response time** during the tests, quantified in terms of the number of simulation steps where each simulation step is 20 ms. This step length is limited by the computation speed of the deep neural network model, which could be significantly improved with highly likely advances in hardware and software. The RL relays have shown a fast response time as listed in Table 10, the longest delay is 4 simulation steps which corresponds to 80 ms. Moreover, the fault detection time of RL relay is not explicitly correlated with fault current magnitude, and is much faster than the melting curve of typical time-delay fuses under all scenarios. We note that, in practice however, the response time could be limited by the data acquisition rate of measurements instruments. This fast response time also allows an ample time window for additional confidence checks, during which successive flags can be used to reduce false-positives even further.

C. PERFORMANCE: LARGE CIRCUITS - SYNTHETIC AUSTIN FEEDERS

Many real low-voltage distribution feeders have a large circuit with multiple long branches. These large feeders usually have multiple fuses along the feeder lines that coordinate with the recloser at the source to provide additional reliability. The coordination between the recloser and fuses, i.e. the "fuse-saving scheme", consists a vital part of the proper functioning of feeder protection. Faults in distribution systems can be roughly classified as transient or persistent fault. Transients faults (e.g. arc, lighting current surge, animal contact) can be cleared by briefly de-energizing the circuit, a.k.a reclosing, for a very short time and re-connect; persistent faults (e.g. downed conductors, exposed cable) are present regardless of whether the circuit is energized and thus can not be restored by reclosing operations. During faults, it is usually desirable for the substation recloser to quickly operate for a few times to clear transient faults and prevent the fuses from blowing unnecessarily. For reclosers, the quick detection of faults is the key to successful coordination.

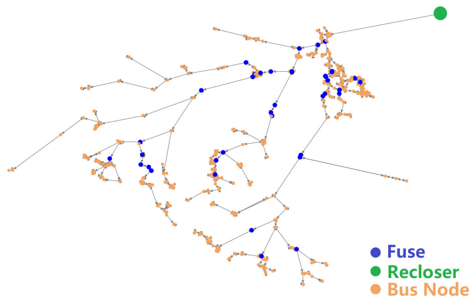


FIGURE 7. **Network Diagram of a Synthetic Austin Feeder (One of the Five Used in the Experiment).**

TABLE 11. **Failure Rate of Relays Under 30% DER in Synthetic Austin Feeders.**

RL Based Relay			
Scenario	False Operation	Occurrences	Probability
No Fault	Trip	0 / 25000	0.00 %
Faulted	No Response	0 / 25000	0.00 %
Faulted	Failed Coordination	61 / 25000	0.24 %
Overcurrent Relay			
Scenario	False Operation	Occurrences	Probability
No Fault	Trip	0 / 25000	0.00 %
Faulted	No Response	2881 / 25000	11.52 %
Faulted	Failed Coordination	2194 / 25000	8.77 %

For faults deep in large and long feeders the impedance between substation and the grounding point can be significant enough to cause the fault current seen by the recloser relatively low. Additional energy sources (DERs) can even cause the current measured at the fuse larger than at the substation. These scenarios can cause problem in mis-coordination if the recloser delay is longer than the fuse melting time. Even for instantaneous recloser configurations it is still possible that the fault current is lower than the pickup current to trigger operation (under-reaching). Under heavy DER penetration these two problems become much more frequent and obvious.

We select five representative feeder circuits from the synthetic Austin dataset and test the performance of the RL recloser relay algorithm. Figure 7 shows the topology and fuse placements of one of the circuits. Each circuit has a recloser placed at the source bus that is controlled based on either RL or overcurrent mechanism. The recloser is expected to respond to faults in the feeder before any fuse in the downstream blows. During training, the RL agent is given a negative reward if its operation is slower than the melting curve of any fuses along the feeder. Similar to the experiment presented above, random fault scenarios generated from Monte-Carlo simulation is used to evaluate the relay operations. Table 11 shows the total **failure rate** summary in the 5 selected feeder circuits, each of which are simulated for 5000 random scenarios.

Similarly, the **robustness** is assessed by varying the load capacity, percentage of DER output and SCMVA. It can be observed in Table 12 that the failure rate of overcurrent relay deteriorates faster in much larger feeders used in

TABLE 12. **Robustness Against Peak Load, DER and SCMVA Increase in Synthetic Austin Feeders.**

Peak Load Increase	+10 %	+20 %	+30 %
RL Failure Rate	0.28 %	0.32 %	0.34 %
Overcurrent Failure Rate	13.2 %	10.49 %	10.12 %
Peak DER Increase	+10 %	+20 %	+30%
RL Failure Rate	0.22 %	0.27 %	0.24 %
Overcurrent Failure Rate	20.69 %	21.16 %	22.52 %
SCMVA Variation	60 %	80 %	120%
RL Failure Rate	0.24 %	0.25 %	0.25 %
Overcurrent Failure Rate	21.58 %	20.55 %	20.03 %

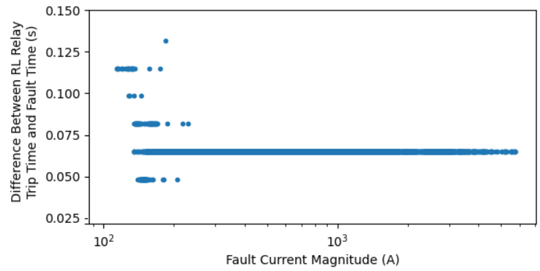


FIGURE 8. **Actual Trip Delay of RL Relay and Fault Current Magnitude.**

this simulation. Note that as the load level increases, the performance of overcurrent improves while RL relay drops slightly. This may be attributed to that most failed episodes for RL relay are false positive or coordination failures, in contrast to all false negatives from overcurrent, which implies RL model in this case is more prone to mis-identify "normal" measurements as faults.

We plot the response speed of RL relay against the observed maximum fault current in all the simulated scenarios in Fig. 8 to illustrate the **response speed**. It can be seen that the speed of RL relay in most scenarios is near constant, except for some cases when the fault current is almost indistinguishable from normal load current. This feature is especially suitable for the protection of large feeder circuits as the fast fault detection leaves a ample time window for coordination reclosing with fuses, while remaining less susceptible to coordination failures compared to instantaneous overcurrent.

V. CONCLUDING REMARKS

This paper introduces and thoroughly tests a deep reinforcement learning based protective relay control strategy for distribution grids with many DERs. It is shown that the proposed algorithm that builds upon existing hardware and uses the same information available to today's overcurrent protection yields faster and more consistent performance. This algorithm can be easily applied in both a standalone relay and in coordination with fuses. The trained RL relays can accurately detect faults under situations where the performance of traditional overcurrent protection deteriorates heavily. The RL relays are robust against unexpected changes in operating conditions of the distribution grid at the time of planning, reducing the need to re-train the relays after deployments. The fast and consistent response speed provides ample time for coordination and breaker operation.

The proposed deep RL relays will be easy to implement with the currently available distribution infrastructure. A particularly attractive feature is that the proposed algorithm for relays can operate in a completely decentralized manner without any communication. This communication-free setting is not only easy to implement for currently available distribution grid infrastructure, but also less vulnerable to potential cyber-attacks. The input to the RL relays are the same as traditional relays so the instrument transformers can be retained during deployment. The training process does not require human intervention since the production of training data and computation of optimal control policy can be fully automated. The weights of the DQN obtained during training can be saved into a general-purpose micro-controller or potentially a more optimized machine learning chip.

In the future, we plan to provide a theoretical guarantee for the convergence of our sequential RL algorithm. We will conduct a thorough and careful investigation of the operation of RL relays under various realistic scenarios by running year-long simulations under variety of stochastic variation of the operational and fault parameters of networks. We will explore potential performance or robustness improvement by using more inputs parameters such as voltage, frequency or apparent impedance. We are working with time domain simulators for more detailed training data generation and fault study with the most realistic models for components including control loops and electromagnetic transients. We will also investigate the possibility of hardware prototyping and Hardware-in-the-Loop test with Real-Time Digital Simulator (RTDS).

REFERENCES

[1] *The Potential Benefits of Distributed Generation and the Rate Related Issues That May Impede Its Expansion*, U.S. DOE, 2007. [Online]. Available: https://www.ferc.gov/sites/default/files/2020-04/1817_study_sep_07.pdf

[2] U. Shahzad *et al.*, "Protection of distributed generation: Challenges and solutions," *Energy Power Eng.*, vol. 9, no. 10, p. 614, 2017.

[3] J. A. Silva, H. B. Funmilayo, and K. L. Bulter-Purry, "Impact of distributed generation on the IEEE 34 node radial test feeder with overcurrent protection," in *Proc. 39th North Amer. Power Symp.*, Sep. 2007, pp. 49–57.

[4] P. K. Dash, S. R. Samantaray, and G. Panda, "Fault classification and section identification of an advanced series-compensated transmission line using support vector machine," *IEEE Trans. Power Del.*, vol. 22, no. 1, pp. 67–73, Jan. 2007.

[5] H. Zhan *et al.*, "Relay protection coordination integrated optimal placement and sizing of distributed generation sources in distribution networks," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 55–65, Jan. 2016.

[6] H.-T. Yang, W.-Y. Chang, and C.-L. Huang, "A new neural networks approach to on-line fault section estimation using information of protective relays and circuit breakers," *IEEE Trans. Power Del.*, vol. 9, no. 1, pp. 220–230, Jan. 1994.

[7] X. Zheng, X. Geng, L. Xie, D. Duan, L. Yang, and S. Cui, "A SVM-based setting of protection relays in distribution systems," in *Proc. IEEE Texas Power Energy Conf. (TPEC)*, Feb. 2018, pp. 1–6.

[8] Y. Zhang, M. D. Ilic, and O. Tonguz, "Application of support vector machine classification to enhanced protection relay logic in electric power grids," in *Proc. Large Eng. Syst. Conf. Power Eng.*, Oct. 2007, pp. 31–38.

[9] H. C. Kilickiran, B. Kekezoglu, and N. G. Paterakis, "Reinforcement learning for optimal protection coordination," in *Proc. Int. Conf. Smart Energy Syst. Technol. (SEST)*, Sep. 2018, pp. 1–6.

[10] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[11] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017.

[12] H. Xu, A. D. Dominguez-Garcia, V. V. Veeravalli, and P. W. Sauer, "Data-driven voltage regulation in radial power distribution systems," *IEEE Trans. Power Syst.*, vol. 35, no. 3, pp. 2133–2143, May 2020.

[13] J. Sun *et al.*, "An integrated critic-actor neural network for reinforcement learning with application of DERs control in grid frequency regulation," *Int. J. Electr. Power Energy Syst.*, vol. 111, pp. 286–299, Oct. 2019.

[14] R. El Helou, D. Kalathil, and L. Xie, "Fully decentralized reinforcement learning-based control of photovoltaics in distribution grids for joint provision of real and reactive power," *IEEE Open Access J. Power Energy*, vol. 8, pp. 175–185, 2021.

[15] M. Bagheri, V. Nurmanova, O. Abedinia, and M. Salay Naderi, "Enhancing power quality in microgrids with a new online control strategy for DSTATCOM using reinforcement learning algorithm," *IEEE Access*, vol. 6, pp. 38986–38996, 2018.

[16] T. P. I. Ahamed, P. S. N. Rao, and P. S. Sastry, "A reinforcement learning approach to automatic generation control," *Electr. Power Syst. Res.*, vol. 63, no. 1, pp. 9–26, Aug. 2002.

[17] D. Wu, X. Zheng, D. Kalathil, and L. Xie, "Nested reinforcement learning based control for protective relays in power distribution systems," in *Proc. IEEE 58th Conf. Decis. Control (CDC)*, Dec. 2019, pp. 1925–1930.

[18] M. Glavic, "(Deep) reinforcement learning for electric power system control and related problems: A short review and perspectives," *Annu. Rev. Control*, vol. 48, pp. 22–35, Jan. 2019.

[19] M. E. Harmon and S. S. Harmon, "Reinforcement learning: A tutorial," Wright Lab., Air Force Mater. Command, Wright Patterson AFB, OH, USA, Tech. Rep. WL-TR-97-1028, 1997. [Online]. Available: https://apps.dtic.mil/sti/pdfs/ADA323194.pdf

[20] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[21] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[22] B. Bakker, "Reinforcement learning with long short-term memory," in *Proc. 14th Int. Conf. Neural Inf. Process. Syst., Natural Synth. (NIPS)*, Vancouver, BC, Canada. Cambridge, MA, USA: MIT Press, 2001, pp. 1475–1482. [Online]. Available: https://dl.acm.org/doi/10.5555/2980539.2980731

[23] S. Kapoor, "Multi-agent reinforcement learning: A report on challenges and approaches," 2018, *arXiv:1807.09427*.

[24] G. Brockman *et al.*, "OpenAI gym," 2016, *arXiv:1606.01540*.

[25] M. Abadi *et al.* (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. [Online]. Available: https://www.tensorflow.org/

[26] K. P. Schneider *et al.*, "Analytic considerations and design basis for the IEEE distribution test feeders," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3181–3188, May 2018.

[27] H. Li *et al.*, "Building highly detailed synthetic electric grid data sets for combined transmission and distribution systems," *IEEE Open Access J. Power Energy*, vol. 7, pp. 478–488, 2020.

[28] G. Ruan *et al.*, "A cross-domain approach to analyzing the short-run impact of COVID-19 on the US electricity sector," *Joule*, vol. 4, no. 11, pp. 2322–2337, Nov. 2020.

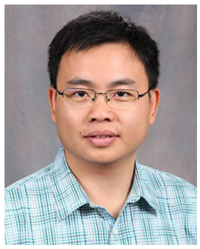
[29] A. Pregelj, M. Begovic, and A. Rohatgi, "Quantitative techniques for analysis of large data sets in renewable distributed generation," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1277–1285, Oct. 2004.

DONGQI WU (Student Member, IEEE) received the B.S. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 2017. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Texas A&M University. He worked at Centerpoint Energy as an Engineer Intern in 2019. His current research interests include data science and machine learning application in power grid resilience enhancement.

DILEEP KALATHIL (Senior Member, IEEE) received the M.Tech. degree from the IIT Madras and the Ph.D. degree from the University of Southern California (USC) in 2014. He was a Post-Doctoral Researcher with the EECS Department, University of California at Berkeley, Berkeley, CA, USA, from 2014 to 2017. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. His research interests include intelligent transportation systems, sustainable energy systems, data driven optimization, online learning, stochastic control, and game theory. He was a recipient of the Best Academic Performance in the EE Department from the IIT Madras and the Best Ph.D. Dissertation Prize in the USC Department of Electrical Engineering.

MIROSLAV M. BEGOVIC (Life Fellow, IEEE) is currently the Department Head of Electrical and Computer Engineering and the College of Engineering Excellence Professor at Texas A&M University. Prior to that, he was a Professor and the Chair of the Electric Energy Research Group, School of Electrical and Computer Engineering, and an Affiliated Faculty Member of the Brook Byers Institute for Sustainable Systems and the University Center of Excellence in Photovoltaic Research, Georgia Institute of Technology. He delivered over 100 keynote and invited presentations worldwide. He was a recipient of the IEEE PES 2019 Meritorious Service Award. He is a Former Chair of the Emerging Technologies Coordinating Committee of IEEE PES, the IEEE PES Treasurer (2010–2011), and the IEEE PES Distinguished Lecturer. He served as the President-Elect, the President, and the Immediate Past President of the IEEE Power and Energy Society (2012–2018).

KEVIN Q. DING (Member, IEEE) received the Ph.D. degree in electrical engineering from Texas A&M University and the M.B.A. degree from the University of Houston. He has been with CenterPoint Energy since 2001 and his responsibilities include implementation, deployment and operation support of transmission energy management system (EMS), advanced distribution management system (ADMS), and advanced metering system (AMS). He is currently an Advanced Digital Systems Director of the Operational Technology and Cyber Security Division, CenterPoint Energy. He is a licensed Professional Engineer in the State of Texas and a Certified Information Systems Security Professional (CISSP).



LE XIE (Fellow, IEEE) received the B.E. degree in electrical engineering from Tsinghua University, Beijing, China, in 2004, the M.S. degree in engineering sciences from Harvard University, Cambridge, MA, USA, in 2005, and the Ph.D. degree from the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA, in 2009. He is currently a Professor with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. His research interests include modeling and control of large-scale complex systems, smart grids application with renewable energy resources, and electricity markets.