Reinforcement Learning based Proactive Control for Enabling Power Grid Resilience to Wildfire

Salah Uddin Kadir, *Student Member, IEEE,* Subir Majumder, *Member, IEEE,* Anurag Srivastava, *Fellow, IEEE,* Ajay Chhokra, *Member, IEEE,* Himanshu Neema, Abhishek Dubey, *Senior Member, IEEE,* and Aron Laszka

***Abstract*—Industrial electric power grid operation subject to an extreme event requires decision-making by human operators under stressful conditions. Decision making using system data informatics under adverse dynamic events, especially if forecasted, should be supplemented by intelligent proactive control. Power transmission system operation during wildfires requires resiliency-driven proactive control for load shedding, line switching, and resource allocation considering the dynamics of the wildfire and failure propagation to minimize the impact on the system. However, the possible number of line and load switching in an extensive industrial system during an event make traditional prediction-driven and stochastic approaches computationally intractable, leading operators to often use pre-planned or greedy algorithms. In this work, we model and solve the proactive control problem as a Markov decision process and introduce an integrated testbed for spatio-temporal wildfire propagation and proactive power-system operation. Our approach allows the controller to provide setpoints for all generation fleets in the power grid. We evaluate our approach utilizing the IEEE test system mapped onto a hypothetical terrain. Our results show that the proposed approach can help the operator to reduce load outage during an extreme event. It reduces power flow through lines that are to be de-energized, and adjusts the load demand by increasing power flow through other lines.**

***Index Terms*—Industrial Power System, Proactive Control, Intelligent Control, Reinforcement Learning, Resiliency, Wildfire.**

I. Introduction

A. Background

CHANGING climate raises the potential for frequent wildfire events with catastrophic consequences on critical industrial systems. To prevent the potential of originating secondary wildfire hazards [1], several utilities in the USA typically restrict power flow through some of their assets during emergency events (e.g., public safety power shut-off, or PSPS, events in the state of California, USA [2]). Compared to other extreme weather events, the slow progression of wildfires [3] across the large geographical span of industrial power systems provides grid operators with sufficient time to proactively control generator set-points in real-time to prevent rolling black-out or even prevent cascading outages (as observed in the 1977 New York black-out [4]). Here, we define proactive control as any pre-event or during-event action to minimize the expected impact of an evolving extreme event.

Traditional power transmission system operation is governed by power system economics and the $N-1$ operational reliability criterion. However, in the advent of increasing frequency of extreme weather events, operators across North-America are thinking about stalling the economics-based operation and move the system into resiliency mode. The operators would rely on the emergency warning and caution (EWAC) direction received from various stakeholders during the transition. Techniques, such as Markovian model-based proactive sequential re-dispatch of generators [5], resilience metric-driven coordinated decision [6], stochastic resource allocation approach [7], integrated proactive control [8], the coordinated control of multiple microgrids connected via transmission system [9], co-planning of transmission lines and distributed resources [10], defensive islanding formation [11], optimal power shut-offs [12], power shut-off with restoration under limited budget [13], fair de-energization scheduling [14], are common approaches for resilient transmission network operation and control. There are also a plethora of tools for optimal proactive control of power distribution systems (see [15], [16] as examples).

B. Challenges

Furthermore, these proactive tools have limited industrial use within the modern-day control center. Traditionally used $N-1$ security constrained criterion utilized for economic load dispatch or optimal power flow also comes under the proactive scheme. The control center operators regularly utilize scenario-based dynamic and static security assessments to determine control actions for each kind of potential impact on the power system, or if the impact materializes, then an emergency action plan, including remedial action schemes, will be needed to be deployed. These scenario-driven corrective control taken after the system impact lacks customizability, and hence they are truly sub-optimal.

In a deregulated environment, during the normal operating mode, or during the emergency condition, some of the control actions may require coordination among various agents, such as generating entities, Independent System Operators (ISOs) / Regional Transmission Operators (RTOs), Distribution System Operators (DSOs), aggregators, and load-serving entities. Abundant monitoring data from supervisory control and data acquisition (SCADA) and phasor measurement units (PMUs) can be leveraged in determining the outage forecast and EWAC. These forecasts can be utilized for proactive

S. U. Kadir is with the University of Houston, Houston, TX. S. Majumder and A. Srivastava are with the School of Electrical Engineering and Computer Science, West Virginia University, Morgantown, WV. (e-mail: anurag.k.srivastava@mail.wvu.edu). A. Chhokra, H. Neema, and A. Dubey are with the Vanderbilt University, Nashville, TN. A. Laszka is with the Pennsylvania State University, State College, PA. Authors acknowledge that this work was in part supported by National Science Foundation (NSF) awards 1840192, 1840083, and 1840052.

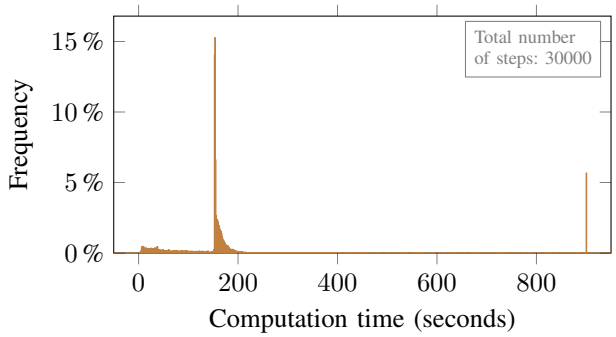


Fig. 1: Computation time distribution of the optimizer with conventional multi-period rolling-horizon optimization approach over 100 episodes.

resource allocation and deployment of necessary measures to serve consumers during emergency conditions. In this regard, statistical models of outage duration [17], [18] can be applied by the operators for proactive resource allocation and measures to serve the customers during emergency condition. However, the involvement of line and generator outages during extreme events with combinatorial nature makes classical optimization-based approaches mixed-integer nonlinear programming (MINLP) in nature, making the control task computationally challenging and resource-intensive to handle in real-time. Discussed forecast-based multi-period rolling-horizon proactive optimal approaches, model-predictive or robust control approaches suffer from computational challenges as demonstrated in Fig. 1, and hence they are difficult to be deployed for real-time applications. As shown in the figure, there is a significant percentage of the simulation not reaching the optimality gap of 0.01% within the computation time of 900 s or 15 minutes (computation time might significantly change depending upon the choice of solvers, but integer linear programming problems are NP-hard in general), demonstrating its lack of deployability. In almost all cases, the increased computational time is needed to demonstrate the optimality of the feasible solution and when the integer values are different compared to the initial solution. Furthermore, recent events have shown the unsuitability of traditional forecast models in an effective determination of outage risk (e.g., majority of the wind power generators were outaged due to lack of winterization in the 2021 Texas power outage event).

Although recent advances in machine learning (ML) have paved its way into power system, its scope has mostly been limited to electric load, price, renewable generation forecasting, fault, and failure analysis [19] and outage predictions [20]. Despite recent literature on using machine learning for operational support from RTE, France [21] and others [22], lack of worst-case guarantee has posed an obstacle to using ML algorithms for power system operational support [23]. In the current context of proactive control of the power system, utilizing historical wildfire propagation data, an ML-based controller can help in real-time control of the power system in the advent of the disaster, alleviate scenario-based emergency action deployment as a corrective measure, resulting in an efficient operation. This also alleviates the difficulties of traditional proactive optimization and rule-based approaches while minimizing the possibility of human errors due to operating in a stressful environment. Therefore, the objective is: *can we leverage recent advancements in ML-based approaches to develop proactive control approaches providing us with decision support for the power transmission system as the extreme weather event progresses?* In this regard, reinforcement learning (RL) based power-system decision support has gained significant traction in recent years [24], [25]. This, along with recent advances in RL-based control [26], indicates the plausibility of successful deployment of ML-based techniques for proactive control in the advent of a disaster. Although the proposed approach provides a decent feasibility guarantee, in case the power system operators are skeptical of the use of 'black-box' in the power system operation, the proposed formulation could be suitably tuned to provide initial points for the conventional rolling-horizon approaches.

C. Contributions

In this paper, a novel approach to aid power-system operators in effective proactive intelligent control of available resources during wildfires has been proposed. The core contributions of this paper are as follows:

- Developed and formulated the proactive control problem as a Markov decision process (MDP) for the power system to minimize load outages considering the time horizon of the wildfire event.
- Proposed a novel approach to solve the proactive control problem, which is an ensemble of a compact representation for the agent's observation, action processing, and a deep reinforcement learning (DRL) based power generation coordination approach.
- Developed an integrated testbed[1] combining the wildfire and the power-system simulator, which captures the impact of the wildfire on the power-system assets.

Due to limited available literature, the efficacy of the proposed approach has been compared against myopic control policies. As discussed later in the paper, the control action is determined solely based on the predicted next step. Given the proposed DRL-based approach has limited foreseeing capability, it can be envisaged that the proposed proactive approach can reduce load losses resulting in reduced outages compared to myopic control policies, which are also robust regarding different fire propagation. The controller is able to provide the decision-support agent with generation control setpoints. The integrated testbed for training and evaluating the performance of the developed agent utilizes an IEEE standard transmission system mapped onto a topographical map with careful consideration of the locations of power transmission sub-stations.

D. Organization

Section II provides the integration of the simulator models that allow MDP formulation. Section III introduces the MDP formulation, and Section IV provides deep reinforcement

[1]The testbed and the agent will be available as open-source software.

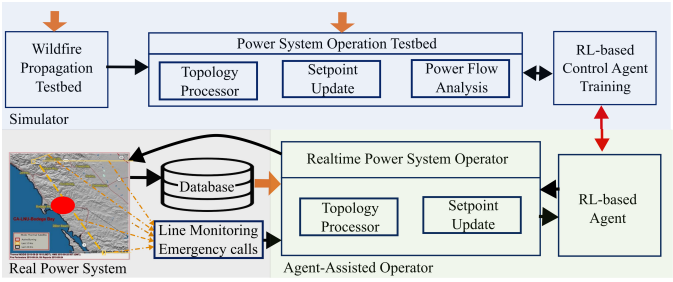


Fig. 2: Overview of the proposed work with integration of wildfire propagation and power system operational response

learning-based agent's training procedure. Section V discusses the simulation setup and the experimental results, and Section VI finishes with concluding remarks.

II. Simulator Models

Fig. 2 shows a high-level overview of the power-system operation supplemented by an external RL agent/controller during the wildfire events. The integrated model comprises two major components: (i) an offline integrated testbed to facilitate training of the RL agent and (ii) online deployment of the trained agent in the natural environment. This section provides a detailed treatment of the integrated testbed comprising of wildfire propagation testbed and power system operation testbed, followed by their real-world equivalent. The list of symbols utilized in Sections II and IV are provided in Table I. The first set of symbols in the table corresponds to wildfire propagation and topography to power system topology mapping; the second set of symbols corresponds to the power system operations; and the third set corresponds to the DRL-based agent control.

A. Wildfire Propagation Model

1) Wildfire Propagation Dynamics in Topographical Space: Stochastic wildfire propagation model as given in [27] has been utilized here to develop the testbed. Entire geographical region comprises of a grid cell set, X. The temporal horizon is also divided into uniform-length contiguous steps of k. The state $s^f_{x,k}$ of fire in each cell $x \in X$ at the beginning of k^{th} interval (i.e., at the k^{th} time step) is captured by (i) status of fire represented by a boolean variable d^x_k, and (ii) available wildfire fuel within the cell represented by an integer variable h^x_k (≥ 0).

The boolean variable d^x_k can be in either one of two states: non-ignited $d^x_k = 0$ and ignited $d^x_k = 1$. Once a cell is ignited, it consumes fuel at a constant rate of C^x until it exhausts (burns) all fuel in the cell. Precise fuel availability dynamics is given in (1). Once all the fuel is burnt, the cell returns to the non-ignited state. Associated set of equations for ignition dynamics are omitted for brevity.

$$h^x_k = \begin{cases} h^x_{k-1} & \text{if} \quad \neg d^x_{k-1} \vee h^x_{k-1} \leq 0 \\ h^x_{k-1} - \mathrm{C}^x & \text{otherwise.} \end{cases} \tag{1}$$

The evolution of status of fire d^x_k is stochastic and driven by the transition probability ρ^x_k. Specifically, the probability

Symbol	Description
X, x	Grid, cell $x \in X$
M	Total number of cells
k, Δk	Time-step, step-size
s^f_k	State of wildfire at k
d^x_k	Burning status of cell x at k
C^x	Fuel burning rate of cell x
h^x_k	Amount of fuel in cell x at k
ρ^x_k	Probability of cell x being ignited at k
$\mathrm{P}^y_{x,k}$	Probability of fire spreading from cell y to x
$\mathcal{H}^x_k$	Neighboring cells of x that can contribute to fire spread to x
i,N	Node i, Set of nodes or buses $i \in N$
t, T	Transmission line t, Set of branches or transmission lines $t \in T$
z^f_i, z^f_t	Operational status of substations, and transmission lines
$\mathcal{L}$	Labeling function
$G_{(\cdot)}$	Symbolizes a set of cells corresponding to a given power system asset
s^p_k	State of power system at k
z^o_i, z^o_t	Operational state of substations/transmission lines due to operator action
z^e_i, z^e_t	Operational state of equipments due to external input
$w^{c\_l}_i$, $w^{nc\_l}_i$	Weights associated with critical and non-critical load at node i
$P^g_{i,k}$	Power generation output at node i at time step k
$\Delta P^{c\_l}_i$, $\Delta P^{nc\_l}_i$	Critical and non-critical load removed from node i
Φ_{k+1}	Set of decision variables of the operator
ϵ	Minuscule positive bias
$\theta_{i,k}$	Voltage angle at node i at time step k
$P^{flow}_{t,k}$	Power flowing through line t at time step k
$z^g_{i,k+1}$	Current operating state of node i
P^{min}_i, P^{max}_i	Minimum and maximum power generation output at node i
R^{max}_i	Maximum ramp rate of generating station at node i
v^e_i	Select node i for the power adjustment
ΔP^e_i	incremental setpoint adjustment (ramp)
Γ^0	Very large constant
$\mathrm{P}^l_{i,k}$	Available load demand at node i in time step k
α_i	Fraction of critical load at node i
B_t	Susceptance of line t
θ^{min}_i, θ^{max}_i	Minimum and maximum values of voltage angles at node i
$\mathcal{D}$	Markov decision problem
S, A	State and action space for $\mathcal{D}$
$\mathcal{R}$	Reward function for $\mathcal{D}$
μ	Actor policy
γ	Discount factor
N^{sld}	Maximum servable load demand
S_f	Fire simulator
$S_{pm}, S_{pr}, S_{pr'}$	Power simulator (Myopic transition), power simulator (RL transition), Power simulator (Myopic-assist)
$a_k, or \mu(s)$	actor-network generated values
a^{rl}_k	RL agent's action
r^m_k, r^{rl}_k	Myopic reward, RL reward at time-step k
r^{rm}_k	RL-transition based power simulator's Myopic reward at time-step k

TABLE I: List of frequently used notations

of cell x being ignited at k^{th} time step (i.e., $d_k^x = 1$) is given by (2).

$$\rho_k^x = \begin{cases} 0 & \text{if} \quad \neg d_{k-1}^x \wedge |\mathcal{H}_k^x| = 0 \\ 1 - \prod_{y \in \mathcal{H}_k^x} (1 - \mathrm{P}_{x,k}^y) & \text{if} \quad \neg d_{k-1}^x \wedge |\mathcal{H}_k^x| > 0 \\ 1 & \text{otherwise.} \end{cases} \quad (2)$$

Here, $\mathrm{P}_{x,k}^y$ denotes the probability of fire spreading from cell y to x. For a given cell x, $\mathcal{H}_k^x$ is the set of topographical neighboring cells that can contribute to the spreading of fire ($d_k^y = 1$) to cell x in k^{th} time step. The state of the whole topographical grid s_k^f can be obtained by composing the state of every cell, shown in (3).

$$s_k^f = \left\langle (d_k^1, h_k^1), \cdots, (d_k^M, h_k^M) \right\rangle \quad (3)$$

where $(1, ..., M)$ are M cells in the grid.

2) Topography to Topology Mapping: To facilitate understanding the impact of propagating wildfire on the power system operation, the power system environment is geotagged with the cell information. Note that the entire state of wildfire is not necessary to capture this interaction, and a reduced set, $\mathcal{L}(s_k^f) = \left\langle d_k^1, \ldots, d_k^M \right\rangle$, is used in this regard. Topologically, the power system can be represented as a collection of nodes, N, representing generation and transmission substations, connected through a set of transmission lines, T. At a given time step, for each node $i \in N$ and branch $t \in T$, given cell-level fire propagation status obtained earlier, binary variables $z_{i,k}^f$ and $z_{t,k}^f$ indicate the operational status of substations and transmission lines respectively and is captured using (4).

$$z_{(\cdot),k}^f = \begin{cases} 0 & \text{if} \quad \exists x \in G_{(\cdot)} \text{ s.t. } \mathcal{L}(s_k^x) \neq \text{non-ignited} \\ 1 & \text{otherwise} \end{cases} \quad (4)$$

Here set $G_{(\cdot)}$ symbolizes the cells corresponding to a given power system asset. Nodes corresponding to the same substations, underground cables will also required to be appropriately represented.

B. Power System Operation Model

Typically, an existing energy management system (EMS) provides decision-support to the operator in solving traditional multi-period scenario-based deterministic or stochastic operational optimization problem for the decision making. However, as shown in Fig. 2, the system operator is a myopic entity that actively seeks the recommendation of the RL-agent in the decision-making in the wake of the disaster. The operator also monitors the system state along with the system-wide emergency condition. Since the applicability of the proposed controller is limited to the wildfire time horizon, and the operator is expected to return to economic mode following expiration emergency conditions, the operating horizon of the controller is finite. Operator actions can be divided into two stages:

1) Topology Revision: It has been considered that the action taken by the operator is based on the observation made at the beginning of k^{th} interval, or at k^{th} time step. Here, the operator receives the emergency responses from the wildfire propagation testbed, given by $z_{i,k}^e$ and $z_{t,k}^e$ (see (4)), and first deploys them faithfully. Consequently, the revised operational status of a substation and transmission line for the $k+1^{th}$ time step becomes $z_{i,k+1}^o = z_{i,k}^e z_{i,k}^o$ and $z_{t,k+1}^o = z_{t,k}^e z_{t,k}^o$ respectively (where $= 1$ denotes availability, and $= 0$ denotes shut-off condition of the transmission assets).

2) Setpoint Update: Given the updated system topology ($z_{i,k+1}^o$ and $z_{t,k+1}^o$), existing power system state (s_k^p), and partial generator control setpoints from the controller at k^{th} time step, myopic operator determines the complete list of setpoints of the generators and load shedding schedule while aiming to minimize the value of the lost load at $k+1^{th}$ time step. As indicated earlier, contrary to traditional economics-driven power system operation, it has been considered that the operator's objective should be to minimize the value of the shedded load at $k+1^{th}$ time step[2], and such an objective will remain in effect until the emergency condition is lifted. The consequent problem to be solved is given as:

$$\min_{\Phi_{k+1}} \sum_{i \in N} w_i^{c\_l} \Delta P_{i,k+1}^{c\_l} + w_i^{nc\_l} \Delta P_{i,k+1}^{nc\_l} + \epsilon \left(P_{i,k+1}^g - P_{i,k}^g \right)^2 \quad (5)$$

where Φ_{k+1} consists of the set of decision variables of the operator that includes power generation outputs $P_{i,k+1}^g$ (without any loss of generality, all the generators at a given bus are aggregated), operational state of generators group $z_{i,k+1}^g$, critical, non-critical loads shedded ($\Delta P_{i,k+1}^{c\_l}$, $\Delta P_{i,k+1}^{nc\_l}$), nodal angles $\theta_{i,k+1}$, of node $i \in N$ along with line flows, $P_{t,k+1}^{flow}$, through branch $t \in T$. Also, $w_i^{c\_l}$, $w_i^{nc\_l}$ (> 0) are the values of critical and non-critical loads, respectively. The criticality of the loads (hospitals, fire stations, police stations, etc., are generally treated as critical loads) imposes the condition of $w_i^{c\_l} \geq w_i^{nc\_l} > 0$ $(\forall i \in N)$. Here, we assume that the generators are equipped with load rejection capabilities [28], and thusly, the entirety of the generating fleet could be brought offline, and consequent solution is a feasible solution. Therefore, (5) is always expected to provide a feasible solution. Overall state of the power system is captured by $s_{k+1}^p = \{\Phi_{k+1} \cup z_{t,k+1}^o \cup z_{i,k+1}^o\}$.

Sole utilization of the value of load loss expression as an objective may lead to the existence of multiple feasible solutions with undesirable ramping of the generators. Addition of an expression $\sum_{i \in N} \left(P_{i,k+1}^g - P_{i,k}^g \right)^2$ in the objective with minuscule positive bias of ϵ inhibits such possibility, with little to no impact on the original objective. Furthermore, $\min w_i^{nc\_l} \gg \epsilon > 0$. The objective function must be subject to the following power system operation and safety constraints:

a) Generation Constraints: Finite generation capacity (lower and upper limits given by P_i^{min}, P_i^{max}, respectively) and ramping capabilities (given by R_i^{max}) limit the operability of the generators. Here, $z_{i,k+1}^o$, and $z_{t,k+1}^o$, as discussed earlier will provide nodal and line availability, respectively, which will be utilized to revise operating limit of the generating fleet. Furthermore, given the generators can suffer from forced outages, the generator statuses are tracked using $z_{i,k}^g$.

[2]The operator may also use a multi-period greedy algorithm to derive and deploy set points for next time step.

As a part of the load rejection capability[3], when generators face forced outages, their (down) ramping rate (see (8)) and operating limit can be allowed to contravene (see (7)), subject to the outaged generators will not be brought online without a thorough safety check. For this paper, these generators will remain outaged indefinitely (see (6)). To model such a condition, the help of a large positive real constant, Γ^0, as shown in (8) is sought. These conditions, along with revised generating and ramping capability, are given in the following equations. Here, Δk represents the power system operating interval.

$$0 \leq z^g_{i,k+1} \leq z^g_{i,k} z^o_{i,k+1} \tag{6}$$

$$z^g_{i,k+1} \mathrm{P}^{min,*}_i \leq P^g_{i,k} \leq z^g_{i,k+1} \mathrm{P}^{max,*}_i \tag{7}$$

$$-\Delta k R^{max}_i \leq P^g_{i,k+1} - P^g_{i,k} \leq \Delta k R^{max}_i + \Gamma^0 \left(1 - z^g_{i,k+1}\right) \tag{8}$$

b) Load Demand Constraints: The necessity of the deployment of control action for generators at k^{th} time step to ensure load-generation balance at $(k+1)^{th}$ time step requires prediction of load demand. Available historical data, shown in Fig. 2 (using brown arrows), can facilitate such computation. However, the development of load prediction models is beyond the scope of this paper and assumed to be given.

Limited availability of generation during the prevailing contingencies necessitates demand curtailment. Suppose, $\mathrm{P}^l_{i,k+1}$ is the operator predicted load demand, then, following removal of associated substation, updated load demand will be $z^o_{i,k+1}\mathrm{P}^l_{i,k+1}$. Additionally, α_i is the parameter representing the critical load fraction (positive real number) served. Availability of a large number of switchable loads within the distribution network connected at the transmission substation enables treating the sheddable loads as a continuous variable [29]. Consequently, critical and non-critical sheddable loads $(\Delta P^{c\_l}_{i,k+1}, \Delta P^{nc\_l}_{i,k+1})$ is bounded as follows:

$$0 \leq \Delta P^{c\_l}_{i,k+1} \leq \alpha_i z^o_{i,k+1} \mathrm{P}^l_{i,k+1} \tag{9}$$

$$0 \leq \Delta P^{nc\_l}_{i,k+1} \leq (1-\alpha_i) z^o_{i,k+1} \mathrm{P}^l_{i,k+1} \tag{10}$$

c) Load Flow Constraints: Since the typical operating voltage within the power system remains close to 1.00 *pu*, and the difference in the voltage angle of the adjacent buses is tiny, we consider a DC power flow model [30]. An associated mathematical expression is given in (12). Here, B_t is the element corresponding to the t^{th} branch in the imaginary part of the nodal admittance matrix. Equation (11) represents the nodal flow balance equation. Also, the set $T^i \subseteq T$ consists of all the branches that are connected to node i. Here, θ^{min} and θ^{max} are upper and lower bound of nodal angle, respectively. The power flow constraint is described in (14).

$$P^g_{i,k+1} - \mathrm{P}^l_{i,k+1} z^o_{i,k+1} + \Delta P^{c\_l}_{i,k+1} + \Delta P^{nc\_l}_{i,k+1} - \sum_{t \in T^i} P^{flow}_{t,k+1} = 0 \tag{11}$$

$$P^{flow}_{t,k+1} - z^o_{t,k+1} \mathrm{B}_t \left(\theta_{i,k+1} - \theta_{j,k+1}\right) = 0 \tag{12}$$

$$\theta^{min} \leq \theta_{i,k+1} \leq \theta^{max} \tag{13}$$

$$-z^o_{t,k+1} \mathrm{P}^{maxflow}_t \leq P^{flow}_{t,k+1} \leq z^o_{t,k+1} \mathrm{P}^{maxflow}_t \tag{14}$$

3) Power Flow Analysis: At the beginning of each time step, the operator waits for the duration of Δk to calculate the revised setpoints. A power flow analysis tool[4] has been utilized to calculate slack-bus set-point in an effort to calculate system-wide state. In actual deployment, it has been assumed that the slack-bus generators are equipped with AGCs to ensure load-generation balance, and as shown in the online deployment part of Fig. 2, the correct system state can be directly obtained from the real environment. It is notable that the delay in the measurement of the state and deployment of the control action is minuscule enough to account for.

III. MDP Formulation

In this proposed model, the power system operator is myopic. As a result, the controller needs to consider future trajectory of the wildfire propagation and provide an appropriate control signal to the operator, facilitating prevention from running into reliability related issues while maximizing the value of load served. Given the probabilistic nature of wildfire propagation and power system loads, the control problem is formulated as a Markov decision process (MDP) problem. An MDP is a tuple $\mathcal{D} = \langle S, A, \mathcal{P}, \mathcal{R} \rangle$, where S is a finite state space, A is a finite action space, $\mathcal{P}$ is the transition probability function and $\mathcal{R}$ is the reward function. The agent chooses an action from the possible action space to lead the system from one state to another. For the given problem, these elements are defined as follows:

States: The state $s^{\mathcal{E}}$ of the environment $\mathcal{E}$ includes the state of the fire model and the state of the power system $s^{\mathcal{E}} = \langle \mathcal{L}(s^f), s^p \rangle$. Where, $\mathcal{L}(s^f) = \{0,1\}^M$ is the fire status of each cell. The s^p includes the status of each power system component $\{0,1\}^{|T|+|N|}$, the current set points of the generators $\mathbb{R}^{|N^{gen}|}$, and the load demand of each node $\mathbb{R}^{|N|}$. The complete state space S is defined in (15).

$$S = \underbrace{\{0,1\}^M}_{\mathcal{L}(s^f)\,\text{(wildfire)}} \times \underbrace{\{0,1\}^{|T|+|N|} \times \mathbb{R}^{|N^{gen}|} \times \mathbb{R}^{|N|}}_{s^p\,\text{(power system)}} \tag{15}$$

Actions: The agent adjusts the generation set points of each generator based on the current state. So, the action space is $A = \mathbb{R}^{|N^{gen}|}$, where $N^{gen} \subseteq N$ is the set of nodes with generation capabilities.

Transitions: An MDP evolves as a result of the set of actions taken. The transition probability function, $\mathcal{P}(s_{k+1}|s_k, a_k)$ indicates that the action a_k at time step k in state s_k will lead to the next state s_{k+1}. Here, the stochasticity arises from the wildfire propagation model, changing load demand and shutting off power system components because of fire.

[3]Generators are isolated from the bulk power system into a load-bank. So, the generators become immediately invisible to the bulk power system operators.

[4]AC power flow equations can be invoked here.

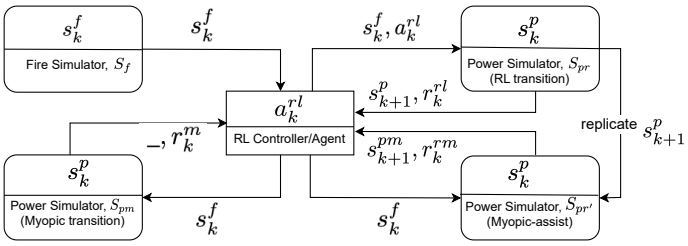


Fig. 3: DRL-based agent's learning model.

Reward Function: The reward function, $r_k = \mathcal{R}(s_k, a_k, s_{k+1})$, indicates that the reward is received for taking action a_k in state s_k to reach next state s_{k+1}.

$$r_k = -\sum_{i \in N} \left((1 - z_{i,k}^o)\mathrm{P}_{i,k}^l + \Delta P_{i,k}^{c\_l} + \Delta P_{i,k}^{nc\_l} \right) \tag{16}$$

In (16), the expression $(1 - z_{i,k}^o)\mathrm{P}_{i,k}^l$ signifies the shedded load demand following isolation of a substation, and $\Delta P_{i,k}^{c\_l} + \Delta P_{i,k}^{nc\_l}$ denotes aggregated critical and non-critical load curtailment of substation i.

Policy: A policy function $\mu(s)$ specifies the action a to be taken in state s. At every time step k, the agent selects an action, $a_k = \mu(s_k)$, based on the deployed policy. This experiment aims to find an optimal policy that can maximize the cumulative sum of expected rewards in each episode. The optimal policy in the model are defined in (17), where n is the finite number of steps in each episode.

$$\mu^* = \arg\max_\mu \mathbb{E}\left[\sum_{k=0}^{n} \mathcal{R}(s_k, a_k, s_{k+1}) \right] \tag{17}$$

IV. Deep Reinforcement Learning based Proactive Control

A standard reinforcement learning-based approach is considered to train the agent. The agent learns by iteratively updating its policy μ^*, defined in (18), at every step k, where $\gamma \in (0, 1)$ is a temporal discount factor for infinite-horizon future rewards.

$$\mu^* = \arg\max_\mu \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k \cdot \mathcal{R}(s_k, a_k, s_{k+1}) \right] \tag{18}$$

As discussed earlier, a finite number of steps in an episode is considered in the test environment as the episode finishes at some point. In contrast, an infinite horizon with a temporal discount factor, a standard RL approach, is considered for training.

1) Training procedure overview: Fig. 3 shows the simulator's state transition in each step. The Testbed includes one fire simulator, S_f, and two power simulators, e.g., the Power simulator (myopic transition), S_{pm}, and the Power simulator (RL transition), S_{pr}. Both power simulators are functionally the same; the power simulator S_{pm} transition happens based on the fire impact on the power system, and the power simulator S_{pr} transition happens based on the fire impact as well as the RL actions. The Testbed also creates another instance, e.g., the Power simulator (myopic-assist), $S_{pr'}$, of S_{pr} in every step after taking an RL action. Note that the power simulator S_{pm} is needed to calculate the *custom reward* and the power simulator $S_{pr'}$ is needed to calculate the *servable load demand*, described in the later sections, to train the RL controller.

The RL controller collects fire information s_k^f at time step k from the fire simulator S_f. The controller first observes the Myopic transition based on the current fire impact on the power simulator S_{pm} and collects the Myopic reward r_k^m. Then the controller again observes the Myopic transition on the replica power simulator instance $S_{pr'}$ based on the current fire impact and collects the next state power system information s_{k+1}^{pm} and the RL-transition based power simulator S_{pr}'s Myopic reward r_k^{rm}. Finally, the controller takes RL action a_k^{rl} on the power simulator S_{pr}, collects the RL reward r_k^{rl} and the next state power system information s_{k+1}^p. At this point, the Testbed creates a replica instance of power simulator S_{pr} at state s_{k+1}^p, e.g., $S_{pr'}$ to use it in the next state.

Note that the Myopic transition happens based on the impact of the fire on the power system, and then the internal operator adjusts the set points if a bus or branch is removed because of the fire. The RL controller does not do anything in the case of the Myopic transition but uses the transition information, e.g., the reward and the next state information, to train itself. Once the RL controller is trained, it only acts on the Power simulator(RL transition) and Power simulator(myopic-assist) of the deployed environment.

A. Processing State Information

As described in Section III, the dimension of the complete state space is $S = \{0,1\}^{M+|T|+|N|} \times \mathbb{R}^{|N|} \times \mathbb{R}^{|N^{gen}|}$. In practice, the number of cells M is far greater than the number of substations $|N|$ and transmission lines $|T|$ (i.e., $M \gg |N \cup T|$). So, instead of using the fire states of each cell, we convert it into a "fire-distance metric" for each component $N \cup T$. It calculates and observes the geographical distance from the nearest ignited cell. This transformation reduces the wildfire state space from $\{0,1\}^M$ to $\mathbb{R}^{|T|+|N|}$. The reduced state space $\hat{S}$ of the agent is given in (19):

$$\hat{S} = \{0,1\}^{|T|+|N|} \times \mathbb{R}^{|N^{gen}|+2|N|+|T|} \tag{19}$$

Actually, the power system operator does not need to be concerned about all the fire distances. It needs to take care of only those fires that are close to the power system component. Additionally, the learning can be ineffective if all the fire distances from each component are fed to the neural network. So, the operator only considers the fire distances that are apart for a specific number of cells from the power system components. If the considerable fire distance is Y, then the converted fire distance $\hat{d}$ from each component $N \cup T$ is defined in (20).

$$\hat{d} = \begin{cases} 1 - d/Y & \text{if} \quad d < Y \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

Here, $\hat{d} = 0$ means the component is safe for now, $\hat{d} = 1$ means that the fire has already reached the component, and

other components are vulnerable to the fire; whereas components that are closer to 1 are highly vulnerable. This conversion makes fire distances highly sensitive to the neural network and helps fast learning. Now, the power system component status needs not to be used separately. Additionally, the load demand of a bus is constant throughout the whole episode if it does not get fire. As $\hat{d} = 1$ of a bus means the bus is shut off, the load demand also needs not to be used as an input. The final state space S that is fed to the neural network to train the agent is defined in (21).

$$S = \mathbb{R}^{|N^{gen}|+|N|+|T|} \tag{21}$$

Other input data, e.g., the generator's current set point, are also converted into the $[0, 1]$ range.

B. Processing Action

1) Challenges: The power system has different constraints, e.g., generation constraints, load demand constraints, and load flow constraints described in Section II-B. The RL agent needs to take action that satisfies all the constraints. But, the untrained neural network outputs random actions at the beginning, and the agent also adds noises in training for exploration. Now, if there is an excess power generation that cannot be served, it will shut off the pertained generators to satisfy the constraints or incur load losses if there is lower power generation. Once a generator gets shut off, it does not turn on for the rest of the episode. It limits the exploration area and makes it hard to learn. Again, the experiences are not very helpful for learning if the action violates the constraints. The most valuable experiences are those experiences in which the taken action switches the power generation of generators but does not incur any load loss or excess power generation.

2) Total Maximum Servable Load Demand: Myopic control, described in Section V-B1, always takes the best action based on the current power system state without violating constraints but not considering the future. So, the agent can use the Myopic control approach to calculate the maximum possible power generation that can be served based on the current state of the environment, satisfying all the constrained; this is called *maximum servable load demand*, N^{sld}, throughout this paper. If the RL action generates more power than the total N^{sld}, it will violate the constraints and shut off the generator. So, the RL agent must ensure the action never generates more power than the total N^{sld}. In Fig. 3, the agent takes a Myopic action on the Power simulator(myopic-assist), which returns the generation set points as next state information, s^{pm}_{k+1}. The sum of these Myopic generation set points is the total maximum servable load demand.

3) Normalization: As generating a lower amount of power incurs load losses, the RL agent also tries to generate power closer to the total N^{sld}. As described in section V-A, the actor-network uses the "softmax" activation function, which outputs values for all the generators that sum up to 1. The actor-network suggested values, $\mu(s)$, is transformed to the actor-suggested generator's set point, a^r, by multiplying them with the total N^{sld}, defined in (22).

$$a^r = \mu(s) \times \sum N^{sld} \tag{22}$$

But, there is a high chance that this actor-network suggested generator's set point a^r values will violate the ramp limit or other constraints. So, the agent calculates each generator's lower, l_i, and upper bound, u_i, based on generators current set point P^g, max ramp value R^{max}, and maximum/minimum power generation (P^{max}/P^{min}), defined in (23) and (24).

$$l_i = \max\left\{P^g_i - R^{max}_i, P^{min}_i\right\} \tag{23}$$

$$u_i = \min\left\{P^g_i + R^{max}_i, P^{max}_i\right\} \tag{24}$$

Now, the agent normalizes the actor-network suggested generator's output a^r based on the lower and upper bound (26) while ensuring that the total generation will be close to the total N^{sld}, but never exceeds (27). [5]

$$a = \text{argmin}_a \sum_i (a_i - a^r_i)^2 \tag{25}$$

$$\text{such that} \quad \forall i: \; l_i \le a_i \le u_i \tag{26}$$

$$\sum_i^{N^{gen}} a_i = \sum_i^{N^{gen}} N^{sld}_i \tag{27}$$

In this procedure, the RL agent switches power generation from one generator to another but maintains the total power generation closer to total N^{sld}. This approach does not guarantee to serve all the generated power all the time because of load flow constraints. But, it definitely reduces the constraints violations significantly.

4) Connected Components: The whole power system grid can be disconnected into multiple parts because of fires. In that case, the total N^{sld} and the generator's set point need to be calculated for each connected component separately based on the actor-network outputs.

C. Custom Reward for Training

The reward is a negative value of incurred load losses described in Section III. But, if a bus is shut off because of fire, it will incur load losses at every step for the rest of the episode, but that load losses are unrecoverable and not helpful for learning. So, a custom reward is calculated for the training based on the Myopic transition of the environment. As mentioned earlier, in Fig. 3, the Power simulator (myopic-transition) transition happens based on the impact of the fires, whereas the Power simulator (RL transition) transition happens based on RL action with the same fire impact. Based on the returned rewards, we calculate the custom reward, $r^{rl}_k - r^m_k$, to train the agent. The custom reward also incentivizes the RL agent to do better than the Myopic transition.

D. Training Procedure

Algorithm 1 shows the step-by-step sequence to train the DRL agent. At the beginning of each episode, the Testbed resets all of its simulators, initializing each and returning the initial state. The transition of the current state s_k to the next state s_{k+1} happens when it takes action through the $step()$ method. The agent does not need the Myopic state information, so the symbol '_' was used for simplicity.

[5] We used the Python scipy.optimize.minimize() method for this conversion.

Algorithm 1 Training procedure

1: **while** *non-converge* **do**
2: $s_k^f \leftarrow S_f.reset()$
3: $\_ \leftarrow S_{pm}.reset()$
4: $s_k^p \leftarrow S_{pr}.reset()$
5: $S_{pr'} \leftarrow S_{pr}$
6: $s_k \leftarrow$ *pre_processing*(s_k^f, s_k^p)
7: **while** $k \leq max\_step$ **do**
8: $\_, r_k^m \leftarrow S_{pm}.step(s_k^f)$
9: $s_{k+1}^{pm}, r_k^{rm} \leftarrow S_{pr'}.step(s_k^f)$
10: $a_k \leftarrow$ *agent.actor*(s_k) ▷ e.g. $\mu(s)$
11: $a_k^{rl} \leftarrow$ *post_processing*(a_k, s_{k+1}^{pm})
12: $s_{k+1}^p, r_k^{rl} \leftarrow S_{pr}.step(s_k^f, a_k^{rl})$
13: $S_{pr'} \leftarrow S_{pr}$
14: $r_k \leftarrow r_k^{rl} - r_k^m$
15: $s_{k+1}^f \leftarrow S_f.step()$
16: $s_{k+1} \leftarrow$ *pre_processing*(s_{k+1}^f, s_{k+1}^p)
17: *replay_buffer*$.add\_record(s_k, a_k, r_k, s_{k+1})$
18: *agent*$.DDPG($*replay_buffer*$.get\_batch())$
19: **end while**
20: **end while**

The agent preprocesses state information in lines 6 and 16 to feed it into the actor neural networks. In lines 8 and 9, the agent observes the myopic transition of the power simulators S_{pm} and $S_{pr'}$ and collects the rewards and the next state information. The next state information s_{k+1}^{pm} is used to determine the total N^{sld}. In line 10, the agent feeds the preprocessed state information s_k to the actor neural network to generate the actor-network suggested action a_k. In line 11, the agent calculates the RL action a_k^{rl} based on the actor-network generated values a_k and the total N^{sld}. In line 12, the agent takes an RL action on the power simulator S_{pr} and collects the next state information and the RL reward. In line 13, the Testbed creates a replica instance of the power simulator S_{pr}. In line 14, the agent calculates the custom reward based on the myopic reward to train the neural network. The agent stores the current experience (line 17) and uses them to train the actor and the critic networks utilizing the Deep Deterministic Policy Gradient (DDPG) [31] algorithm. Interested readers are referred to the associated paper for the details about the algorithm, as the algorithm has been closely followed to train the agent.

V. Simulation Results

A. Simulation Setup

A standard IEEE 24-bus reliability test system (RTS), superimposed on a geospatial terrain, divided into a 350×350 grid, has been considered here for analysis. Power system operational parameters can be obtained from [32]. Thereupon, there are ten controllable power generators. The entire power system control horizon is divided into time steps with a 5-minute interval, where each episode consists of 300-time steps corresponding to approximately one day.

For simplicity, the load demand within the power network is considered to be deterministic in nature with constant magnitude. As discussed, network-wide loads are comprised of critical and non-critical fractions. This fraction also remains constant throughout the network. The proposed controller tracks spatio-temporal wildfire propagation and provides set points for all generators. Successively, the power system operator would solve the optimization problem to calculate and deploy the complete set of requisite control actions based on load forecasts. To determine the set points for the operator, the SCIP solver in *General Algebraic Modeling System (GAMS)* is utilized due to its versatility.

Fig. 4 shows the simulation progress of an example episode with an interval of 50 steps. The west coast was chosen as the geospatial terrain for the testbed. Each cell is categorized based on the vegetation information (e.g., water, deep desert, desert, low vegetation, land, and forest) and the amount of fuel is defined respectively. Each episode starts with a random wildfire origin within the geographical area. The fire spreads over each step based on the predefined fire propagation probability described in Section II-A1. A new wildfire may also originate at a random time step at a random place. For effective training, we set a boundary for the fire origins inside the grid to ensure that the wildfire impacts the power system assets.

The RL-based agent uses deep neural networks with two hidden layers, consisting of 512 and 512 neurons for both the actor and the critic networks. The actor-network uses *rectified linear* activation function in the hidden layers and *softmax* activation function in the output layer. The *softmax* activation function ensures the total power generation is equal to the total maximum servable load demand (described in Section IV-B2). The critic-network also uses the *rectified linear* activation function in the hidden layers, but a *linear* activation in the output layer. The actor and critic learning rates are 0.001 and 0.002, respectively. The agent uses 0.005 to update the target network and 0.9 as the discount factor to calculate the expected return. The training period for the RL agent was ≈ 14 days. The conventional multi-period optimization, described in V-B3, uses an optimality gap of 0.00 or a maximum of 900 seconds to calculate the set points for a single step.

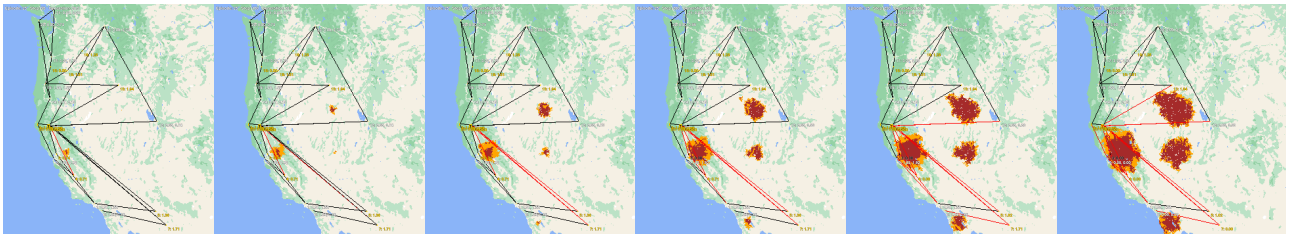
B. Control Approaches

In this experiment, three different control approaches are considered, which are defined as follows:

1) Myopic control: As described in Section II-B2, the operator observes the impact of wildfire on the power system, estimates system wide-load demand, and considers de-energization decision support from the controller to determine the control input for the next time step.

2) Proactive control: In proactive control, the RL-based external controller observes the progress of the fire and provides the set points for the entire generator fleet to the operator. The operator is also expected to observe the impact of wildfire on the power system, estimate system wide-load demand, and account for external control input to calculate and deploy the requisite control actions.

3) Conventional multi-period optimization: Typically power system operators solve rolling horizon optimization



(a) Step 50 (b) Step 100 (c) Step 150 (d) Step 200 (e) Step 250 (f) Step 300

Fig. 4: Simulation progress of an example episode with an interval of 50 steps.

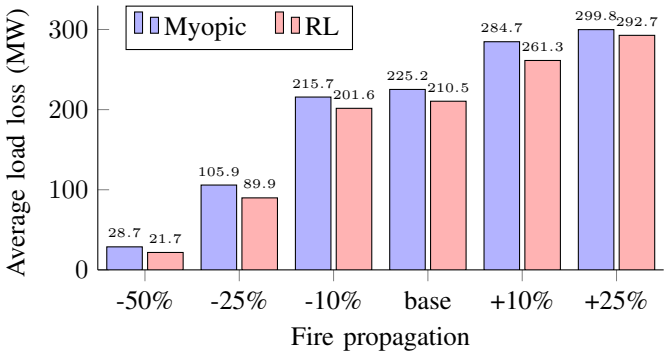


Fig. 5: Average episodic load loss (MW) over 100 example episodes for different fire propagation.

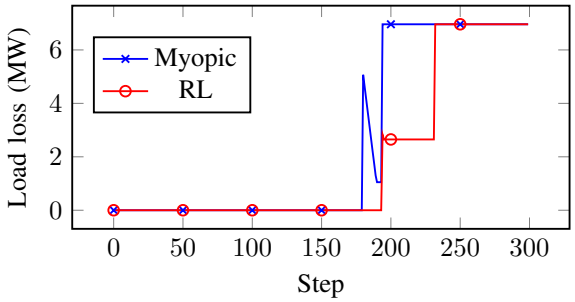


Fig. 6: Step-by-step load losses over an example episode shows the improvement of the RL agent's action.

problems along with real-time load-generation balance. For the rolling horizon problem, the set points of the immediate time step are generally enforceable, while the set points for subsequent time steps are generally used as an advisory signal. In this regard, the operator significantly relies on the forecast of fire propagation for decision-making. Under the same paradigm, a pre-calculated original fire propagation dataset is provided to the conventional multi-period optimizer as forecast data. The Optimizer looks ahead at the required number of time steps of fire propagation to calculate the generation set points.

C. Results

Control approaches	Total load loss (MW)	Worst 5% each step computation time (seconds)
RL	21,048	≈ 0.136
Myopic	22,518	≈ 0.128
Conventional multi-period optimization	16,500	≥ 900

TABLE II: Test results have been calculated using 100 episodes. Here, the conventional multi-period optimization test result is based on perfect fire prediction (using a pre-calculated dataset), which is unrealistic. Still, we calculate this to find the theoretically best possible result given the fire impact on the power system.

Table II shows that the RL agent does substantially better in reducing the load losses than the Myopic agent. Although Conventional multi-period optimization has lower load losses, it takes a long time to converge in critical situations. The external operator cannot bear this long time. The computation time for the RL agent is slightly higher than the Myopic control approach. However, the computation time for both agents is negligible to make suggestions to the external operator as each step is equivalent to 5 minutes interval. So, the higher computation time for the RL agent is not significant here.

Fire propagation can be different based on different geographical dynamics. Fig. 5 shows that the trained agent can take good actions regardless of different fire propagation. The RL agent does not require retraining for differently propagated fire. On the other hand, the RL agent needed higher fire propagation for effective training, as a lower-propagated fire may not interact with the power system in some episodes. The test result indicates that the performance is even better with a lower fire propagation than the originally trained fire propagation. While the improvement is around 6.5% with the same trained fire propagation, it is around 24.4% with 50% less fire propagation and 15.1% with 25% less fire propagation. The RL agent also does substantially better if the fire propagation is even higher than the originally trained fire propagation. The figure also shows that lower fire propagation has lower total load losses as the impact on the power system is also lower.

Figure 6 shows how the RL agent's action reduces the load losses over the Myopic action. The fire removes a branch at step 180, which imbalances power generation and load demand for the Myopic control at some regions of the power system. The RL agent adjusted the power generation, proactively predicting the line removal based on the fire progress. The power system for the Myopic control reduces the load loss over the next couple of steps by adjusting the power generations based on ramp limits. At step 193, a separate fire removes a node that adds load losses for both controllers. Note that

the generator of that node generates more power than the load demand of that node. The RL agent adjusted the power generation proactively, so the load loss was only the load demand of that node. But, for the Myopic control, the load loss was the total power generation. At step 231, the first fire removes another branch that separates a region from the power system. The figure also shows the weakness of the Myopic control approach and the scope of the improvement for a proactive control approach.

VI. Conclusions

This work developed a deep reinforcement learning (DRL) based proactive intelligent control to supplement decision support for industrial power grid operators given a wildfire event. The testbed has been developed by integrating a wildfire-propagation model with a power-system operation model to train and validate a controller that can supplement traditional computationally-intensive, forecast-driven power-system operations during a wildfire. The control problem is formulated as a Markov decision process. The innovative compact representation of observations and actions processing ensures efficient training. Numerical results indicate that the DRL-based proactive control agent can reduce the load loss, which is also robust regarding different fire propagation.

The computationally inefficient conventional multi-period optimization test result using perfect fire forecast is better than the RL test result, if enough time is available. So, theoretically, it is possible to have better results than presented result in this work using some other RL approaches, which will be our future work. We believe that this RL-based approach will spur innovative research in applying AI in the power system to help the industrial operator make decisions during a disaster in timely manner.

References

[1] Western Area Power Lines, "Trees and power lines." [Online]. Available: https://www.wapa.gov/newsroom/FactSheets/Pages/trees-powerlines.aspx

[2] J. T. Abatzoglou, C. M. Smith, D. L. Swain, T. Ptak, and C. A. Kolden, "Population exposure to pre-emptive de-energization aimed at averting wildfires in Northern California," *Environmental Research Letters*, vol. 15, no. 9, p. 094046, 2020.

[3] R. Arghandeh, B. Uzunoglo, S. D'arco, and E. Erman Ozguven, "Guest editorial: Data-driven reliable and resilient energy system against disasters," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, 2022.

[4] J. Latson, "Why the 1977 blackout was one of New York's darkest hours," *Time*, 2015.

[5] C. Wang, Y. Hou, F. Qiu, S. Lei, and K. Liu, "Resilience enhancement with sequentially proactive operation strategies," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2847–2857, 2016.

[6] G. Kandaperumal, S. Pandey, and A. Srivastava, "AWR: Anticipate, withstand, and recover resilience metric for operational and planning decision support in electric distribution system," in *IEEE Transactions on Smart Grid*. IEEE, 2022, pp. 179–190.

[7] M. Movahednia, A. Kargarian, C. Ozdemir, and S. Hagen, "Power grid resilience enhancement via protecting electrical substations against flood hazards: A stochastic framework," in *IEEE Transactions on Industrial Informatics*, 2022.

[8] M. Nazemi and P. Dehghanian, "Powering through wildfires: An integrated solution for enhanced safety and resilience in power grids," in *IEEE Transactions on Industry Applications*, vol. 58, no. 3, 2022.

[9] K. P. Schneider, F. K. Tuffner, M. A. Elizondo, C. Liu, Y. Xu, S. Backhaus, and D. Ton, "Enabling resiliency operations across multiple microgrids with grid friendly appliance controllers," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4755–4764, 2018.

[10] M. Moradi-Sepahvand, T. Amraee, and S. Sadeghi Gougher, "Deep learning-based hurricane resilient co-planning of transmission lines, battery energy storages and wind farms," *IEEE Transactions on Industrial Informatics*, 2022.

[11] M. Panteli, D. N. Trakas, P. Mancarella, and N. D. Hatziargyriou, "Boosting the power grid resilience to extreme weather events using defensive islanding," *IEEE Trans. on Smart Grid*, vol. 7, no. 6, 2016.

[12] N. Rhodes, L. Ntaimo, and L. Roald, "Balancing wildfire risk and power outages through optimized power shut-offs," *IEEE Transactions on Power Systems*, 2020.

[13] N. Rhodes and L. Roald, "Co-optimization of power line shutoff and restoration under high wildfire ignition risk," 2022. [Online]. Available: https://arxiv.org/abs/2204.02507

[14] A. Kody, A. West, and D. K. Molzahn, "Sharing the Load: Considering Fairness in De-energization Scheduling to Mitigate Wildfire Ignition Risk using Rolling Optimization," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, 2022, pp. 5705–5712.

[15] M. Dabbaghjamanesh, S. Senemmar, and J. Zhang, "Resilient distribution networks considering mobile marine microgrids: A synergistic network approach," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5742–5750, 2020.

[16] S. Majumder, G. Kandaperumal, S. Pandey, A. K. Srivastava, and C. Koplin, "Pre-Event Two-Stage Proactive Control for Enhanced Distribution System Resiliency," *IEEE Access*, vol. 10, pp. 83 281–83 296, 2022.

[17] H. Liu, R. A. Davidson, D. V. Rosowsky, and J. R. Stedinger, "Negative binomial regression of electric power outages in hurricanes," *Journal of Infrastructure Systems*, vol. 11, no. 4, pp. 258–267, 2005.

[18] S. D. Guikema, R. Nateghi, S. M. Quiring, A. Staid, A. C. Reilly, and M. Gao, "Predicting hurricane power outages to support storm response planning," *IEEE Access*, vol. 2, pp. 1364–1373, 2014.

[19] N. Kishore, A. Srivastava, and H. Pota, "Guest editorial: Special section on "deep learning and data analytics to support the smart grid operation with renewable energy," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, pp. 6935–6938, 2021.

[20] R. Nateghi, S. D. Guikema, and S. M. Quiring, "Forecasting hurricane-induced power outage durations," *Natural Hazards*, vol. 74, no. 3, pp. 1795–1811, 2014.

[21] B. Donnot, I. Guyon, M. Schoenauer, P. Panciatici, and A. Marot, "Introducing machine learning for power system operation support," *arXiv preprint arXiv:1709.09527*, 2017.

[22] J. Xie, I. Alvarez-Fernandez, and W. Sun, "A review of machine learning applications in power system resilience," in *2020 IEEE Power Energy Society General Meeting (PESGM)*, 2020, pp. 1–5.

[23] S. Chatzivasileiadis, A. Venzke, J. Stiasny, and G. Misyris, "Machine learning in power systems: Is it time to trust it?" *IEEE Power and Energy Magazine*, vol. 20, no. 3, pp. 32–41, 2022.

[24] X. Chen, G. Qu, Y. Tang, S. Low, and N. Li, "Reinforcement learning for decision-making and control in power systems: Tutorial, review, and vision," *arXiv preprint arXiv:2102.01168*, 2021.

[25] M. M. Hosseini and M. Parvania, "Resilient operation of distribution grids using deep reinforcement learning," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 2100–2109, 2022.

[26] P. P. Khargonekar and M. A. Dahleh, "Advancing systems and control research in the era of ML and AI," *Annual Reviews in Control*, vol. 45, pp. 1 – 4, 2018.

[27] D. Bertsimas, J. D. Griffith, V. Gupta, M. J. Kochenderfer, and V. V. Mišić, "A comparison of Monte Carlo tree search and rolling horizon optimization for large-scale dynamic resource allocation problems," *European Journal of Operational Research*, vol. 263, no. 2, 12 2017.

[28] W. P. Gorzegno and P. V. Guido, "Load rejection capability for large steam generators," *IEEE Trans. Power Apparatus and Systems*, vol. PAS-102, no. 3, pp. 548–557, 1983.

[29] M. Esfahani, N. Amjady, B. Bagheri, and N. D. Hatziargyriou, "Robust resiliency-oriented operation of active distribution networks considering windstorms," *IEEE Transactions on Power Systems*, vol. 35, no. 5, 2020.

[30] W. Stevenson, "Element of power system analysis," *McGraw-Hill*, 1975.

[31] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[32] Probability Methods Subcommittee, "IEEE reliability test system," *IEEE Trans. Power Apparatus and Systems*, vol. PAS-98, no. 6, 1979.

Salah Uddin Kadir is currently pursuing a Ph.D. in Computer Science at the University of Houston, Texas, having previously obtained a Master's degree from the same institution. In addition, he has worked as a Software Engineer between 2014 and 2018, following his completion of a Bachelor's degree in Computer Science and Engineering. Salah Uddin Kadir's primary research focus is in the area of automated learning systems and optimization, utilizing techniques such as Reinforcement Learning and Machine Learning.

Subir Majumder (S'17–M'21) received the Ph.D. degree under a Cotutelle/Joint Agreement between Indian Institute of Technology Bombay, India and the University of Wollongong, Australia in 2020.

From 2020 to 2021, he worked as a post-doctoral research associate at Washington State University, Pullman, WA, USA. He is currently working as an Engineering Scientist at the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, USA. He was conferred POSOCO Power System Awards (PPSA) under the Doctoral category in 2020. His research interests include power systems modeling, operations (including operational resiliency) and planning, power system economics, distributed optimization, and the smart grid.

Anurag K. Srivastava (F'22) received the Ph.D. degree in power engineering from the Illinois Institute of Technology, Chicago, IL, USA, in 2005.

He is a Raymond J. Lane Professor and Chairperson with the Computer Science and Electrical Engineering Department West Virginia University. He is also an Adjunct Professor with the Washington State University and Senior Scientist with the Pacific Northwest National Lab. He is an Author of more than 300 technical publications, including a book on power system security and four patents. His research interest includes data-driven algorithms for power system operation and control, including resiliency analysis.

Prof. Srivastava is serving as Chair of PES voltage stability working group, and Vice-Chair of power system operation sub-committee, and Vice-Chair of tools for power grid resilience task force.

Himanshu Neema is a Research Assistant Professor of Computer Science at Vanderbilt University. He received an MS and PhD in Computer Science from Vanderbilt University. Dr. Neema conducts research in model-based design of Cyber-Physical Systems and their integrated co-simulations with hardware and humans. His other research interests include System-of-Systems, Secure and Resilient Systems, Design Automation, Design Space Exploration, Machine Learning, Constraint Programming, Planning and Scheduling, Smart Cities, and Smart Grids. Dr. Neema has 25 years of experience in research and development and has co-authored more than 75 publications.

Ajay D Chhokra is a software architect in the Architecture and Verification of Intelligent Systems at Siemens. He received his Ph.D. in Electrical Energy from Vanderbilt University. His research interests include failure diagnosis in cyber-physical systems, hardware in the loop real-time simulation of energy systems, performance optimization, and workload placement in resource-constrained devices.

Abhishek Dubey is an Assistant Professor of Electrical Engineering and Computer Science at Vanderbilt University, Senior Research Scientist at the Institute for Software-Integrated Systems and co-lead for the Vanderbilt Initiative for Smart Cities Operations and Research (VISOR). His research interests are in the field of artificial intelligence and distributed computing for cyber-physical systems, and smart and connected communities. He is a senior member of IEEE and published several peer-reviewed articles. His key contributions include a robust software model for building cyber-physical applications, along with spatial and temporal separation among different system components, which guarantees fault isolation. He completed his PhD in Electrical Engineering from Vanderbilt University in 2009. He received his M.S. in Electrical Engineering from Vanderbilt University in August 2005 and completed his undergraduate studies in electrical engineering from the Indian Institute of Technology, Banaras Hindu University, India in May 2001.

Aron Laszka is an Assistant Professor in the College of Information Sciences and Technology at The Pennsylvania State University. Previously, he was an Assistant Professor at the University of Houston ('17–'22), a Research Assistant Professor at Vanderbilt University ('16–'17), and Postdoctoral Scholar at the University of California, Berkeley ('15–'16). His research interests revolve around the applications of artificial intelligence and machine learning in societal-scale cyber-physical systems, including transportation and power systems. His work has been supported by the U.S. National Science Foundation, Department of Energy, Department of Transportation, and other federal and industrial sponsors.