Curriculum Based Reinforcement Learning to Avert Cascading Failures in the Electric Grid

Amarsagar Reddy Ramapuram Matavalam
School of Electrical, Computer and Energy Engineering
Arizona State University
Tempe, AZ 85281
amar.sagar@asu.edu

Kishan Guddanti
Pacific Northwest National Laboratory
Richland, WA 99354
kishan.g@pnnl.gov

Yang Weng
School of Electrical, Computer and Energy Engineering
Arizona State University
Tempe, AZ 85281
yweng2@asu.edu

Abstract

We present an approach to integrate the domain knowledge of the electric power grid operations into reinforcement learning (RL) frameworks for effectively learning RL agents to prevent cascading failures. A curriculum-based approach with reward tuning is incorporated into the training procedure by modifying the environment using the network physics. Our procedure is tested on an actor-critic-based agent on the IEEE 14-bus test system using the RL environment developed by RTE, the French transmission system operator (TSO). We observed that naively training the RL agent without the curriculum approach failed to prevent cascading for most test scenarios, while the curriculum based RL agents succeeded in most test scenarios, illustrating the importance of properly integrating domain knowledge of physical systems for real-world RL applications.

1 Motivation & Introduction

The electric power generation is a major contributor of green house emissions. The trend of electrification of transportation and gas infrastructure [1] will further increase this contribution. Thus, in order to achieve the emission goals of the Paris agreement [2], to transition to clean energy and to mitigate the already escalating global climate change, it is necessary to integrate renewables into the power grids and to utilize the existing grid infrastructure to its maximum capacity, while ensuring grid operations are within limits.

A key operational constraint in the power grid is the current flow limit in a transmission line. If left unattended or an appropriate response from the grid operator is delayed, then overloaded lines will be disconnect (trip) and the current that was flowing in these lines will be re-routed through other lines, which can in turn lead to further line tripping. This sequence of events is called a *cascade* [3]. During times of high demand, cascading can lead to regions in the power grid to be disconnected

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2022.

from generators, leading to a blackout [4] (Appendix A). The introduction of renewables that are inherently uncertain is going to exacerbate this phenomenon and so transmission system operators (TSOs) are actively researching methods to prevent cascades. TSOs prefer an economical and flexible solution like *dynamic topology reconfiguration* over the other solutions like load shedding, peak shaving and curtailment [5-7] that increase operational costs.

A dynamic topology reconfiguration (Appendix B) essentially alters the connectivity of nodes and edges in the power grid through closing and opening switches in many substations (nodes) of the grid. The discrete combinatorial nature of the reconfiguration combined with the non-linear relation between load demand, generation and line flows [8] implies that estimating the topologies to prevent cascading as the load demand varies over the course of a day is still beyond the state-of-the-art.

Contribution. We leverage deep reinforcement learning (RL) methods to learn topology controls that prevent cascading failures. We propose an efficient physics-inspired curriculum based deep reinforcement learning approach and demonstrate the superior performance of this approach over the conventional RL methods on the IEEE-14 bus system. We show that the same deep-RL structure is able to learn with our approach but fails when trained using conventional RL approaches.

1.1 Related Work

[9] proposed an expert system-based approach that incorporates both transmission line switching and bus splitting/merging operations. This expert system-based approach is sufficiently fast but suffers from accuracy issues at times, and also, it cannot account for the impact of an optimal control action over a time horizon [10]. [11] includes the time horizon concept but uses a mixed-integer nonlinear optimization method which takes longer times to solve. [12] proposed to learn curriculum strategies as a part of the overall ML-approach for classification or regression. [13] uses curriculum learning to accelerate training of RL agents. However, it has not been explored for controlling network flows.

2 Dynamic Topology Reconfiguration Formulation

There are two ways to dynamically reconfigure the grid topology: 1) transmission line switching and 2) node splitting/merging using the switches within a node. The objective of a dynamic topology controller for the power grid involves identifying the optimal topology grid configuration that minimizes the total line loading on the power grid, given the load demand and generator supply. The sequential planning problem is shown below:

$$\min_{\tau_1 \ldots \tau_t} \sum_{t=1}^{t=n} \sum_{\forall p \in E} \left(\frac{I_{p,t}}{I_{p,max}} \right), \tag{1}$$

$$sub.\ to: f\left(\tau_t, x_t\right) = 0;\ \forall t = \{1, 2, \cdots, n\}, \tag{2}$$

$$\tau_t \in A(\tau_{t-1}, x_t, \rho_S, \rho_C, \rho_H); \forall t = \{2, \cdots, n\}, \tag{3}$$

The line loading is the sum of the ratio of the current at time t ($I_{p,t}$) and the current thermal limits, $I_{p,max}$, $\forall p \in E$ where E is the set of all transmission lines in the power grid. This ratio is also referred to as the *normalized current in a line*. The aim of the topology controller is to minimize the total line loading on the grid over a time horizon $t = \{1, 2, \ldots, n\}$ (equation (1)) by identifying the optimal topology τ_t for every time step t with transmission line switching and bus splitting/merging actions. (2) represents the non-linear power flow constraint of the power grid with a topology τ_t and state vector x_t. The state vector x_t includes the line currents, bus voltages, load injections and generator injections. (3) represents the constraint between topologies in consecutive time steps and accounts for the line disconnection & cascading. The grid topology τ_t should lie in the allowable set of topologies based on the topology and states at the previous time step $(A(\cdot))$.

The parameters ρ_S, ρ_C & ρ_H are identical for all lines and control the time delay and current levels that a line can withstand before disconnecting. The soft threshold, ρ_S, is the ratio I/I_{max} beyond which thermal disconnection can occur with delay. The consecutive overload limit, ρ_C, determines how many time-steps a line can be continuously in soft overload before it is disconnected. The hard threshold ρ_H, is the ratio I/I_{max} beyond which thermal disconnection will occur instantly.

Identifying the optimal topology at for a *single* operating snapshot consists of solving a large scale non-convex mixed-integer non-linear programming problem [14] and introducing time-coupling further increases the computational complexity. To solve this problem in near-real time for grid operations, we turn to deep RL. The sequential topology/cascading dynamics with varying loads/generation are simulated in an RL environment (PyPOWNET [15]) and a reward based on the objective function is used to train a deep RL agent. The actions of the RL agent are the topology reconfigurations at each step. However, directly applying RL to this problem was not successful which led us to develop the curriculum approach.

3 Physics Inspired Curriculum Learning

Curriculum learning is the idea that neural networks learn a difficult task most effectively when first trained on a simpler version of the problem. Curriculum learning is inspired by how humans learn - initially learning simple concepts before attempting complex tasks [12]. It is a form of transfer learning as the knowledge/insights gained by solving simple tasks are leveraged to solve the more complicated task. A proper curriculum (sequence of tasks with increasing hardness) should be designed to apply this approach for effectively learning grid controllers. Another advantage of this approach is that it is agnostic to the RL architecture used as it leverages features of physical systems to enhance the RL-agent performance/training. However, designing an effective curriculum is not trivial, and a bad curriculum can impede agent learning.

We leverage our understanding of the power grid physics and the impact of the parameters ρ_S, ρ_C & ρ_H on the cascading phenomena to simplify the RL environment and design an appropriate reward function. The default parameters of the environment are ρ_S = 1.0, ρ_C = 3 time steps and ρ_H = 1.5. These parameters imply that the overload counter is triggered when a line current exceeds its I_{max}, and the line will be disconnected if the current remains continuously above the ρ_S limit for 3 steps (ρ_C). If the line current exceeds 1.5 times the rating (ρ_H), then it is immediately disconnected.

ρ_S & ρ_H directly impact the line currents beyond which the cascading is initiated. Hence, increasing these parameters from their default value will increase the cascading threshold and expand the feasible topologies/actions that can be used to prevent cascades, thus simplifying the problem. Similarly, ρ_C impacts the time-steps available for the RL-agent before the line disconnects and increasing ρ_C will also simplify the problem.

The reward function, $r(x_t)$, is also modified with a factor α as shown in (4). A key point to emphasize is that increasing these parameters has no impact on the I_{max} of a line and so the reward of a line will become negative if $I > I_{max}$, even if no cascading occurs. This feature is necessary as the negative rewards guide the RL-agent to learn actions that have less overload and so, the transfer learning will be effective with increasing "hardness" till the original parameter settings are reached. To enforce this property explicitly, we increase the value of α as the training "levels" are increased so that the RL-agent is more penalized for a violation at a stricter level compared to an easier level.

$$r(x_t) = \begin{cases} \sum\limits_{p \in E} R_\alpha\left(\frac{I_{p,t}}{I_{p,max}}\right) & \text{if normal time step} \\ -100 & \text{if terminal time step} \end{cases} \qquad R_\alpha(x) = \begin{cases} (0.95 - x) & x \leq 0.95 \\ \alpha \cdot (0.95 - x) & x > 0.95 \end{cases} \tag{4}$$

An additional reason why the curriculum is useful is that the objective function for simpler problems have a smoother behavior than the harder problems and the stochastic gradient steps are more likely to jump over the local maxima [12]. Thus, initializing from the solution of the simpler problem will help the harder problem to be solved faster through stochastic descent.

4 Numerical Experiments

To demonstrate the proposed method, we use the IEEE 14-bus system environment provided by the PyPOWNET [15] package that is developed by RTE, the french TSO. We compare the performance of a conventional A3C agent and a curriculum based A3C agent (CA3C agent) with the same structure. The curriculum consists of 3 levels with level-1 completely disabling cascading and level-3 being the original RL-environment. Additional details of the training are provided in Appendix C. A greedy non-ML agent referred to as the forecasted power flow (FPF) agent is also used to illustrate the advantage of ML. The FPF agent simulates the reward of each action for a single-step forecast of the demand and generation to find the optimal action. More information can be found in [16], [17].

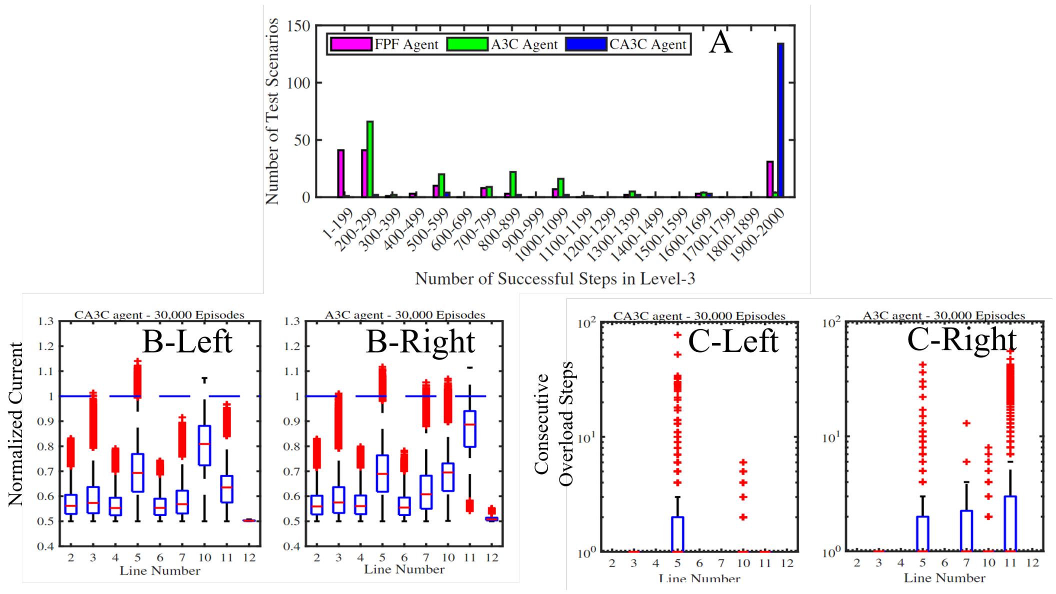


Figure 1: Results demonstrating the improvement in preventing cascading using the curriculum based A3C agent. (Panel-A) The histogram of the number of successful steps for all test scenarios using the FPF, A3C and CA3C agents in level-3 environment. (Panel-B) The box-whisker plots for the normalized current in the critical lines with level-1 environment. (Panel-C) The box-whisker plots for the consecutive overload steps in the critical lines with level-1 environment. [17]

Results. The results are plotted in Figure 1 [17]. We train the A3C and CA3C agents on 50 scenarios and test them on 150 scenarios on level-3 curriculum, with each scenario lasting 2000 time steps. The agents are scored on each scenario based on the number of continuous successful time steps before the scenario terminates due to cascading failure. We also evaluate the A3C and CA3C agents on the test scenarios to illustrate the impact of the curriculum training on the line overloads. We record the normalized line currents for each time-step in each test scenario on level-1 curriculum (cascading disabled). In general, the CA3C agent is much more reliable compared to the A3C agent and the FPF agent in preventing cascading as shown in the histogram on the number of successful time steps for the test scenarios in Panel-A. We also see that the number of time-steps with overload lines in level-1 curriculum is lesser for CA3C agent compared to the A3C agent, as illustrated by the box plots of the normalized current (Panel-B) for the two agents. Finally, The number of consecutive steps with an overload is lesser for the CA3C agent compared to the A3C agent, as illustrated in the box plots in Panel-C. The results in panel-C imply that consecutive overloads $\geq$3 time-steps are more frequent for A3C agent than the CA3C agent in level-1, which leads to more cascading failures in level-3.

Discussion. These results illustrate two key points. First, physically-accurate but greedy approaches (e.g. FPF) that consider the impact of the topology switching for a single time-step in the future are not able to prevent cascading failures due to the inherent time-coupling and memory of the phenomenon. Second, RL agents are flexible enough to learn strategies that prevent cascading, but a nominal RL approach might not be able to learn these agents. Instead, a curriculum based approach that exploits the properties of the power grid can learn RL agents effectively.

5 Conclusion

We introduce the power system thermal cascading problem and demonstrate a physics-inspired curriculum based reinforcement learning approach to address this in an operational setting using topology switching. Our empirical results demonstrate that the proposed approach is superior to conventional deep-RL approaches in learning agents for preventing cascading. The approach can help in identifying topology controls to maintain system security with increasing renewables to reduce carbon emissions and address climate change.

References

[1] E. Zhou, and T. Mai,"Electrification Futures Study: Operational Analysis of U.S. Power Systems with Increased Electrification and Demand-Side Flexibility", 2021, Golden, CO: National Renewable Energy Laboratory. NREL/TP-6A20-79094.

[2] https://www.un.org/en/climatechange/paris-agreement

[3] P. Hines, K. Balasubramaniam, and E. C. Sanchez, "Cascading failures in power grids," IEEE Potentials, 2009.

[4] "Final report on the august 14, 2003 blackout in the united states and canada: causes and recommendations," U.S.-Canada Power System Outage Task Force, Tech. Rep., April 2014. [Online]. Available: http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/BlackoutFinal-Web.pdf

[5] E. B. Fisher, R. P. O'Neill, and M. C. Ferris, "Optimal transmission switching," IEEE Trans. on Power Systems, 2008.

[6] M. Soroush and J. D. Fuller, "Accuracies of optimal transmission switching heuristics based on DCOPF and ACOPF," IEEE Trans. on Power Systems, 2013.

[7] E. Karangelos and P. Panciatici, "'cooperative game'inspired approach for multi-area power system security management taking advantage of grid flexibilities," Phil. Trans. of the Royal Society A, 2021.

[8] D.K. Molzahn and I.A. Hiskens, "A Survey of Relaxations and Approximations of the Power Flow Equations," Foundations and Trends in Electric Energy Systems, vol. 4, no. 1-2, pp. 1-221, February 2019.

[9] A. Marot, B. Donnot et al., "Expert system for topological remedial action discovery in smart grids," IET Digital Library, 2018.

[10] A. Marot, B. Donnot, C. Romero, B. Donon, M. Lerousseau, L. Veyrin-Forrer, and I. Guyon, "Learning to run a power network challenge for training topology controllers," Electric Power Systems Research, 2020.

[11] G. Granelli, M. Montagna, F. Zanellini et al., "Optimal network reconfiguration for congestion management by deterministic and genetic algorithms," Electric Power Systems Research, 2006.

[12] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in Proceedings of the 26th Annual International Conference on Machine Learning. Association for Computing Machinery, 2009.

[13] Y. Wu and Y. Tian, "Training agent for first-person shooter game with actor-critic curriculum learning," in International Conference on Learning Representations, 2017.

[14] F. Pourahmadi, H. Heidarabadi, S. H. Hosseini, and P. Dehghanian, "Dynamic uncertainty set characterization for bulk power grid flexibility assessment," IEEE Systems Journal, 2019.

[15] M. Lerousseau, "Design and implementation of an environment for learning to run a power network (L2RPN)," arXiv preprint arXiv:2104.04080, 2021.

[16] A. R. R. Matavalam, K. P. Guddanti, Y. Weng, and S. Indela, "L2RPN IJCNN 2019 Competition - Second Place Solution." https://github.com/ amar-iastate/L2RPN-using-A3C, 2019

[17] A. R. R. Matavalam, K. P. Guddanti, Y. Weng and V. Ajjarapu, "Curriculum Based Reinforcement Learning of Grid Topology Controllers to Prevent Thermal Cascading," in IEEE Transactions on Power Systems, 2022, doi: 10.1109/TPWRS.2022.3213487.

Appendix A: Thermal Cascading Example

Fig. 2 presents the IEEE 14-bus system with 14 substations, 20 transmission lines and 16 injections (both generations and loads combines). In Fig. 2, the substations are indicated by the nodes (blue circles) in the graph; the yellow circles indicate loads, and the green circles indicate generations. Additionally, as shown in the legend of Fig. 2, each substation has two bus bars, namely "bus 1" and "bus 2". An element (either a line or load or generator) can be located at a substation connected

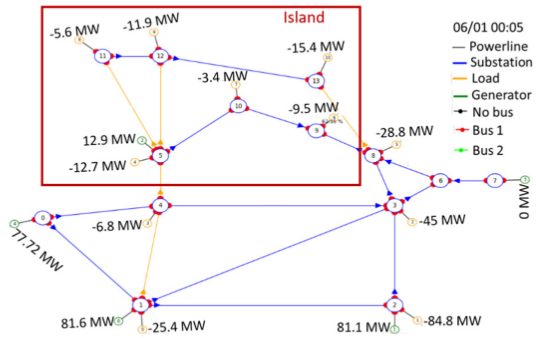


Figure 2: Power grid cascade event in the IEEE 14 bus system due to generation and load injections

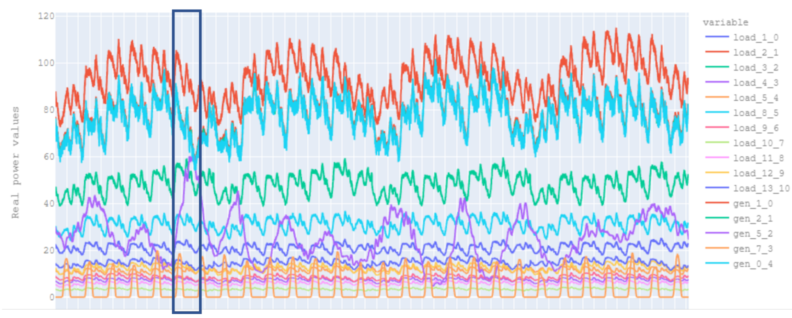


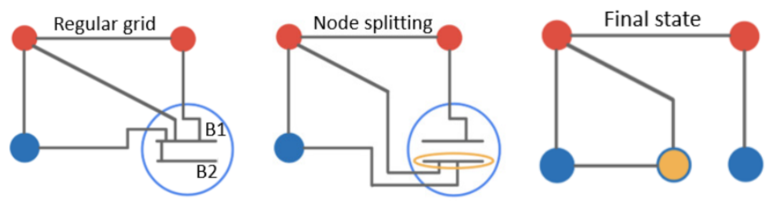
Figure 3: Generation and load injection profiles in the grid versus time for 2000 time steps of 5 minutes. The sharp rise in the wind generation (purple plot) in the boxed region causes a cascading failure.

to either "bus 1" or "bus 2" (node breaker model). To represent the realistic power grid operation scenario, realistic generation and load consumption profiles are injected into the power grid for 2000 time-steps of 5 minutes each (equal to 1 week). Fig. 3 plots the generation and load injection profiles for one scenario.

The injections in Fig. 3 result in a cascading event that leads to power grid blackout. The sudden rise in the wind plant output indicated in the rectangle in Fig 3 is the reason for the cascading event as follows; first, the transmission line connecting substations 1 and 4 are overloaded and becomes out-of-service. The loss of this line reduces the power grid's overall transfer capability, which in turn overloads the other transmission lines in the power grid. This overloading causes the disconnection of the transmission line 4-5 two time steps later. Finally, the transmission lines 8-9 and 8-13 disconnect simultaneously the next time step due to high line loading of 311.72% and 174.11% respectively, resulting in an island as shown in Fig. 2.

Appendix B: Bus Reconfiguration Example

Bus reconfiguration actions are more complex than line switching, and it is explained using a simple 4-bus system from Fig. 4. Fig. 4a presents 4-bus system with five transmission lines and four substations. Each substation in the network has two bus bars to which the power network elements such as loads, generators, transformers, shunt admittances, and transmission lines are connected. Fig. 4a shows a topology with three transmission lines connected to the bus bar 1 (B1) and 2 (B2). For example, as shown in Fig. 4b, a bus splitting action can be triggered to connect two incoming transmission lines to bus bar 2 (B2) and one transmission line to bus bar 1 (B1) separately. This results in a new topology with five nodes, as shown in Fig. 4c, and the new topology can have very different power flow routing properties compared to the original topology.



(a) Power grid with nominal bus bar connectivity.

(b) Bus splitting action using the bus bar connectivity.

(c) Final state of the topology after a bus splitting action.

Figure 4: Bus splitting using bus bars in a substation

Appendix C: Experiment Details

Deep neural networks represent both the actor and critic with two hidden layers of sizes 200 and 50. The first layer of the neural network is shared between the actor and critic leading to joint training of the A3C agent. The learning rate for the actor is 0.0005, and the learning rate for the critic is 0.001. A discount factor equal to 0.95 is used to calculate the time discounted rewards for the training. A total of 50 unique training scenarios are selected from the dataset, and 50 threads are used in parallel during the A3C training procedure. Each unique scenario is made up of 2000 time steps of 5 minutes each that corresponds to 1 week of operation. An agent that continuously operates the grid for all time steps in a scenario is categorized as a successful agent for that scenario. The parameters of level-1, level-2 and level-3 used for the curriculum learning are tabulated in Table-1.

The agents are implemented in Keras and are trained using TensorFlow for 30,000 episodes. The number of successful time steps at each training episode for the two agents is shown in Fig. 5. The median of successful time steps for each of the 30,000 episodes over a window of 15 different scenarios/weeks is plotted in Fig. 5 to smooth out the large variation among the episodes. The enforcement level of CA3C is initially level-1. Based on the agent's performance, the enforcement level is increased to level-2 at episode 6000 and increased to level-3 at episode 14000.

Table 1: Environment parameters for the curriculum levels.

Level	α	ρ_S	ρ_C	ρ_H
1	1	10^9	10^9	10^9
2	5	2	15	10^9
3	10	1	3	1.5

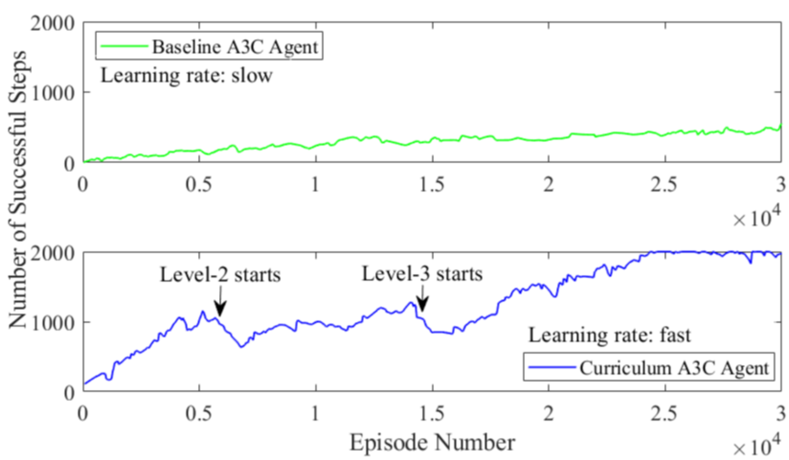


Figure 5: Plot of the number of successful steps versus the episodes during training of A3C and CA3C agents