Efficient Distributed Energy Storage Voltage Control Using Ensemble Deep Reinforcement Learning

James Obert, Rodrigo D. Trevizan, Adrian Chavez
Sandia National Laboratories, Albuquerque, NM

Abstract: To meet the challenges of low-carbon power generation, distributed energy resources (DERs) such as energy storage systems, solar and wind power generators are now widely integrated into the power grid. Because of the autonomous nature of DERs, ensuring properly regulated output voltages of the individual sources to the grid distribution system poses a technical challenge to grid operators. Stochastic, model-free voltage regulations methods such as deep reinforcement learning (DRL) have proven effective in the regulation of DER output voltages; however, deriving an optimal voltage control policy using DRL over a large state space has a large computational time complexity. In this paper we illustrate a computationally efficient method for deriving an optimal voltage control policy using a parallelized DRL ensemble. Additionally, we illustrate the resiliency of the control ensemble when random noise is introduced by a cyber adversary.

Introduction

- Traditional power distribution voltage control use decentralized controls of capacitor banks, on-load tap changers (OLTCs) and smart inverters.
- Physics-based methods for optimization of voltage control require accurate system models.
- Reinforcement learning (RL) is efficient in automated systems where human expertise is not readily available and learns through environmental experiences to make optimal decisions
- Deep learning enabled by artificial neural networks (ANNs) greatly enhanced the RL performance by improving feature extraction and data generalization
- Deep reinforcement learning (DRL) agents can readily learn an optimal actions policy by assessing state-action pairs within an operational environment.
- Current DRL-based autonomous voltage control (AVC) methods use deep Q-learning (DQN) and deep deterministic policy gradient (DDPG) RL but converge very slowly.
- Newer approaches import target Q network and random exploration values during training to achieve faster convergence, but the average reward and controlled voltage variance can be high due to q-value overestimations.

Project Goal:

Develop a computationally efficient, low-variance AVC methodology to modulate reactive power injection of energy storage systems in power distribution systems.

Simulated Grid Environment

- OpenDSS, an electric grid distribution system simulator was used to research the effectiveness and efficiency of the devised AVC algorithm.
- A simulated attack on voltage regulators could cause all voltages in the system to go well below the minimum limit of 0.95 p.u., and possibly causing some equipment to malfunction or not work at all.

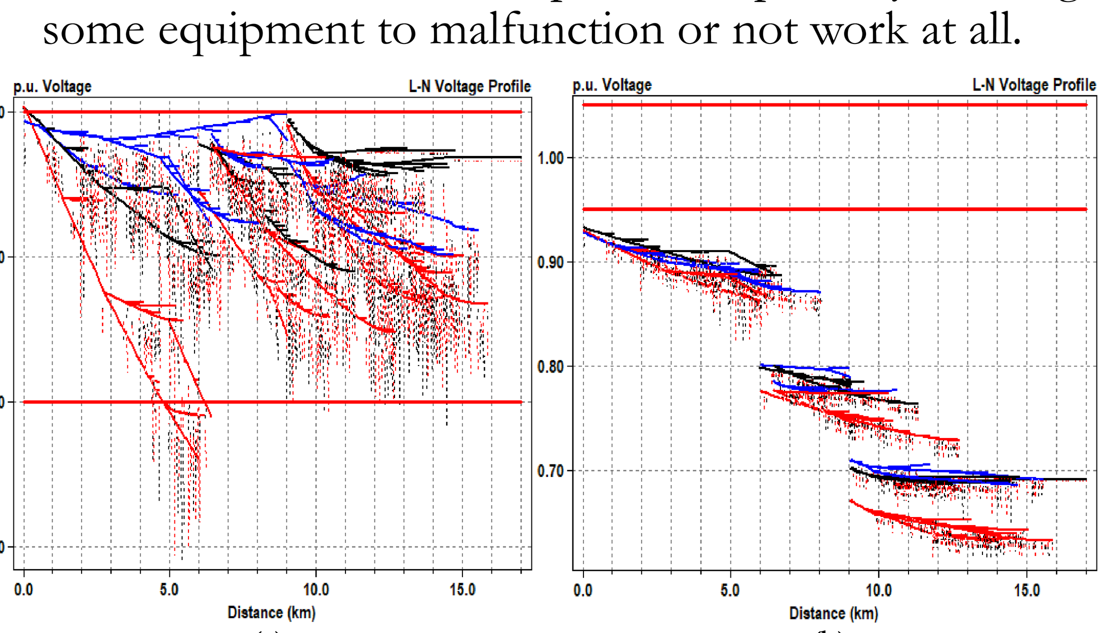

Fig. 1. Simulated voltage profile of a distribution feeder (a) during normal operation, and (b) under a coordinated attack on voltage control devices.

- DER reward function:

$$R_t = \frac{1}{W}\sum_{t'=t-W}^{t}\begin{bmatrix}\frac{k_v}{N_v}\cdot\left(\sum_{j=1}^{N_v} r_{j,t'}^{v}\right)+\frac{k_c}{N_c}\cdot\left(\sum_{j=1}^{N_c} r_{j,t'}^{c}\right)+\\ \frac{k_r}{N_r}\cdot\left(\sum_{j=1}^{N_r} r_{j,t'}^{r}\right)+\frac{k_l}{N_l}\cdot\left(\sum_{j=1}^{N_l} r_{j,t'}^{l}\right)+\\ \frac{k_u}{N_u}\cdot\left(\sum_{j=1}^{N_u} r_{j,t'}^{u}\right)\end{bmatrix}$$

$$r_{j,t}^{v} = \begin{cases} R_{max}-\alpha^{-}\cdot\left(v^{-}-v_{j,t}\right)^2, & v_{j,t} < v^{-}\\ R_{max}, & v^{-}\le v_{j,t}\le v^{+}\\ R_{max}-\alpha^{+}\cdot\left(v_{j,t}-v^{+}\right)^2, & v_{j,t} > v^{+}\end{cases}$$

$$r_{j,t}^{s} = -i_{j,t}^{2}$$

- Energy storage represented by WECC regc a model

DRL Voltage Control

- 2 agents DPG & DQL compose a stacked ensemble.
- Adversary randomly injects values into the inverters.
- The two agents/learners derive an optimal policy using stochastic RL to select their control actions.

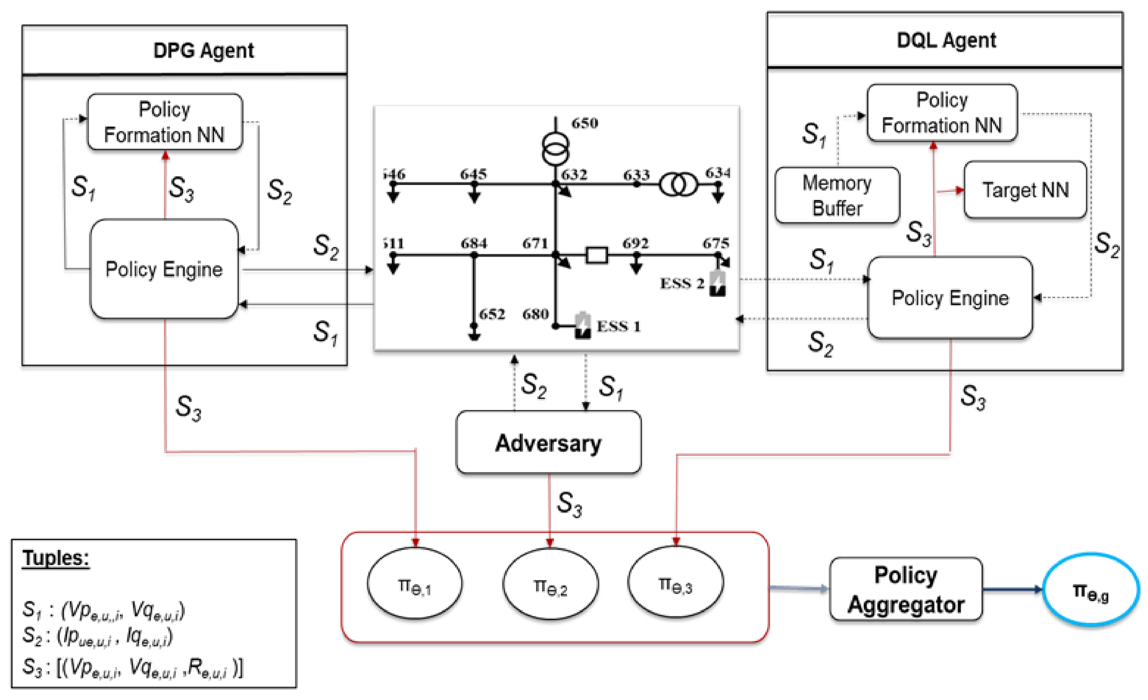

Fig. 2. Ensemble voltage control system.

- DQL & DPG agents utilize deep neural networks to derive and store AVC policies
- Both agents consists of deep ANNs that output a vector of actions stored in separate AVC policies
- DPG & DQL ensemble complement one another
  - DQL predicts future rewards using a value function DPG efficiently observes ANN gradients
- Periodically the DPG & DQL AVC policies are combined into a global optimal policy
- The aggregate AVC policy is used to train a separate deep feed-forward neural net (DFNN) that predicts reward values given newly observed voltages
- Predicting reward values requires less computation using the DFNN than the iterative convergence processes of DQL & DPG agents
- To ensure that the DFNN is trained sufficiently, the DQL & DPG agents run offline in a secondary mirrored simulation environment in parallel with the DFNN and periodically the most recent optimal aggregated policy is used to re-train the DFNN.
- The training process employs a two-layer autoencoding neural net, with its first encoding layer performing linear transformations from input dimension elements to a lower output dimension to ensure a noise-free training dataset.
- The second layer decodes the lower dimension encoding using a linear transformation as well with the output being of the same size as the original input dimension and a sigmoid output layer activation function.

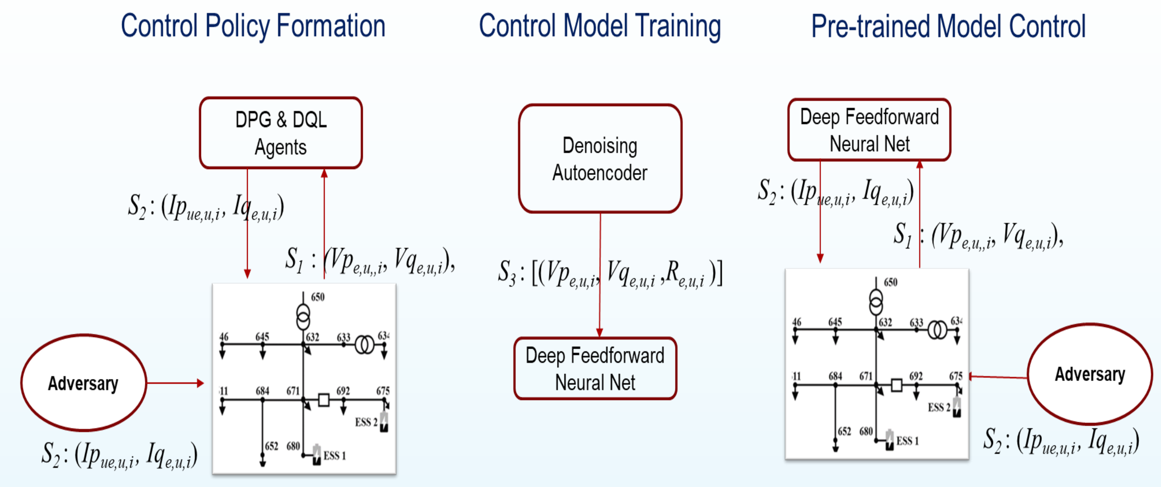

Fig. 3. Voltage control pre-trained neural network.

Case Study

- IEEE 13-bus test feeder with two energy storage systems (ESSs), with load fluctuations
- Intelligent agents perform control of the ESSs to minimize voltage deviations and energy losses
- DRL agents must also mitigate the effects of an adversary that also manipulates setpoints of ESS

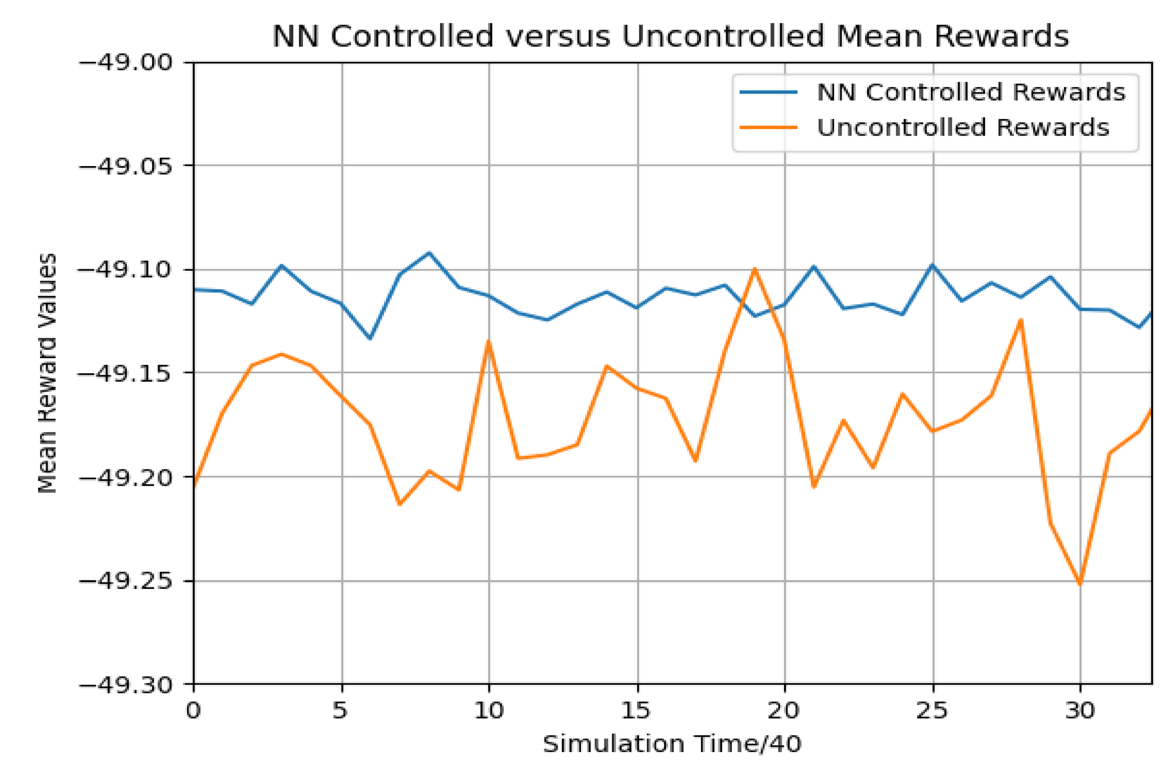

Fig. 4. Neural network controller rewards as simulation progresses.

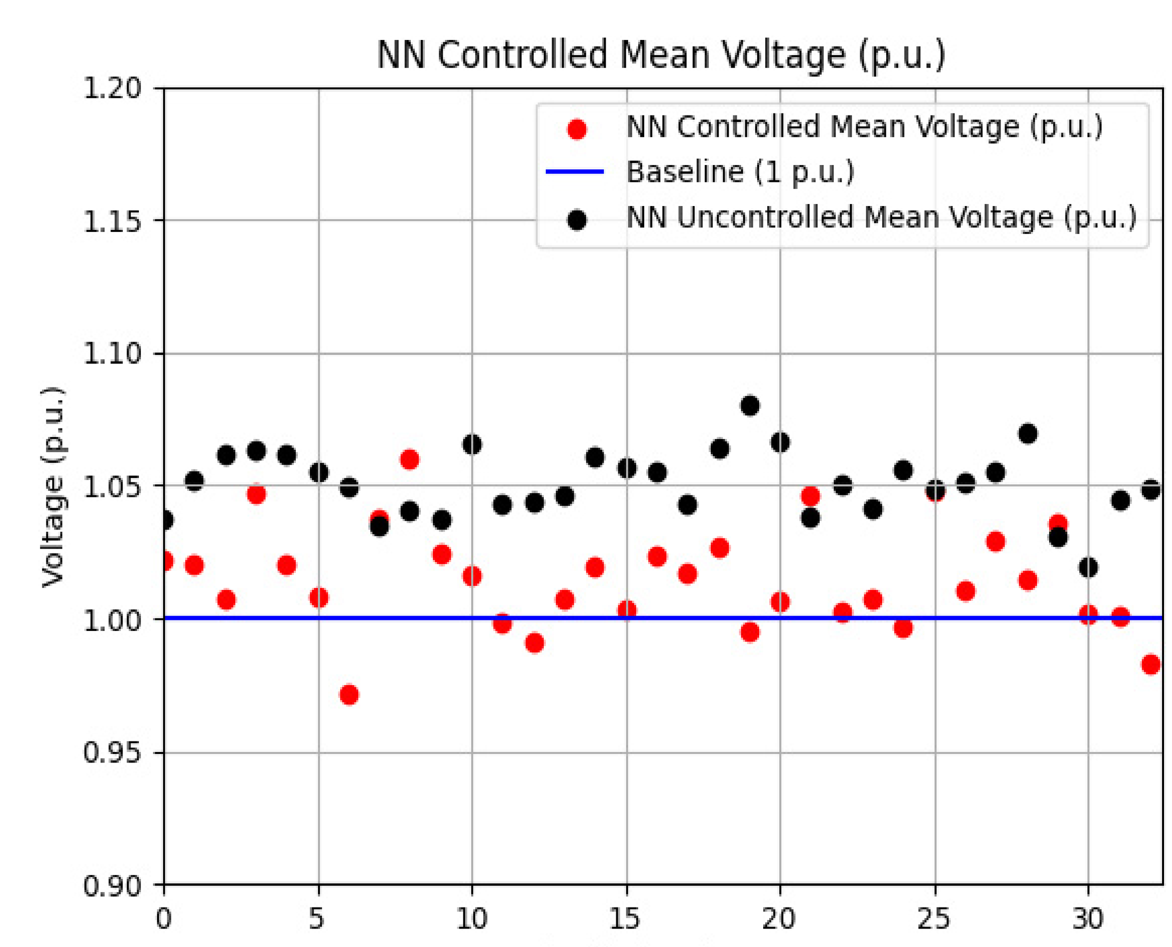

Fig. 5. Neural network- controlled mean voltage (p.u.).

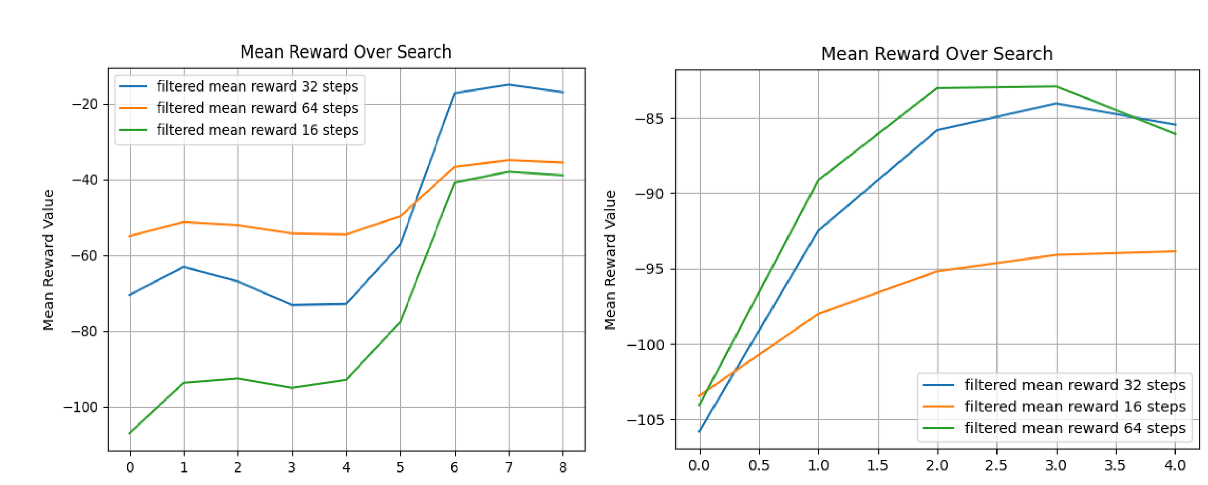

Fig. 6. Mean reward over 5000 training episodes using DPG and DQL.
(a) Mean reward DPG (horzontal axis 5000 episodes/8)
(b) Mean reward DQL (horz. axis 5000 episodes/4)

Conclusion

- DRL ensembles provide stable voltage control for storage-connected inverters
- Stable voltage control is achievable in the presence of noise and cyber adversaries
- By aggregating the control policies of a DRL ensemble and a maximum entropy adversary, it is possible to achieve operational resiliency.
- This solution can be scaled to larger multi-agent control systems using high-performance computing.

Acknowledgement: Funding was provided by the US DOE Energy Storage Program managed by **Dr. Imre Gyuk** of the DOE Office of Electricity Delivery and Energy Reliability
The authors would like to thank Dr. Hyungjin Choi and Dr. Atri Bera for their technical input.