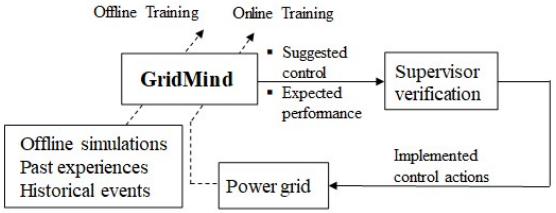
On Training Effective Reinforcement Learning Agents for Real-time Power Grid Operation and Control

Ruisheng Diao[1], Di Shi[1], Bei Zhang[1], Siqi Wang[1], Haifeng Li[2], Chunlei Xu[2], Tu Lan[1], Desong Bian[1], Jiajun Duan[1]
[1]GEIRI North America, San Jose, CA 95134, USA
[2]SGCC Jiangsu Electric Power Company, Nanjing, Jiangsu, China
Contact: Dr. Ruisheng Diao, Ruisheng.Diao@gmail.com, (480)414-7095

NEURAL INFORMATION PROCESSING SYSTEMS

Introduction

- **Grand Challenges**: the increasing dynamics and stochastics in the modern power grid due to increased penetration of renewable energy and power electronics-based equipment make it difficult to design and implement optimal control actions in real time for mitigating operational risks
- **Critical Needs for:**
  - Accurate and fast wide-area monitoring system to detect potential issues
  - Effective optimal control suggestions in real time to support operators
- **Objectives:** to develop effective real-time control strategies to control bus voltages, line flows and transmission losses thin their corresponding limitations before/after disturbance(s)
  - **AI-based Solution -- A measurement-driven, grid-interactive, self-evolving, and open platform for power system automatic dispatch and control.**



Problem Formulation

The mathematical formulation of the control problem is given below:

Objective function :

$$\text{minimize} \sum_{i,j} P_{loss}(i,j), (i,j) \in \Omega_L \quad (1)$$

Subject to :

$$\sum_{n \in G_i} P_n^g - \sum_{n \in D_i} P_m^d - g_i V_i^2 = \sum_{j \in B_i} P_{ij}(y) \quad (2)$$

$$\sum_{n \in G_i} Q_n^g - \sum_{n \in D_i} Q_m^d - b_i V_i^2 = \sum_{j \in B_i} Q_{ij}(y) \quad (3)$$

$$P_{ij}(y) = g_{ij} V_i^2 - V_i V_j (g_{ij} \cos(\theta_i - \theta_j) + b_{ij} \sin(\theta_i - \theta_j)), (i,j) \in \Omega_L; \quad (4)$$

$$Q_{ij}(y) = -V_i^2 (b_{ij0} + b_{ij}) - V_i V_j (g_{ij} \sin(\theta_i - \theta_j) - b_{ij} \cos(\theta_i - \theta_j)), (i,j) \in \Omega_L \quad (5)$$

$$P_n^{\min} \le P_n \le P_n^{\max}, n \in G; \quad (6)$$

$$Q_n^{\min} \le Q_n \le Q_n^{\max}, n \in G \quad (7)$$

$$V_i^{\min} \le V_i \le V_i^{\max}, i \in B \quad (8)$$

$$\sqrt{P_{ij}^2 + Q_{ij}^2} \le S_{ij}^{\max}, (i,j) \in \Omega_L \quad (9)$$

where P_{loss} is power losses on transmission line connecting bus i and bus j; P_n^g is active power injection into bus n; P_m^d is active power consumption at bus m; θ_i and V_i are voltage phase angle and magnitude at bus i. b_{ij} and g_{ij} are conductance and susceptance of transmission line. P_{ij} and Q_{ij} stand for active and reactive power on a transmission line. Eqs. (2)-(5) represent quasi-steady-state conditions of a power grid. Eqs. (6)-(7) are active and reactive power output limits of each generator. Eq. (8) and Eq. (9) specify bus voltage secure zones and line flow limits to be controlled, respectively.

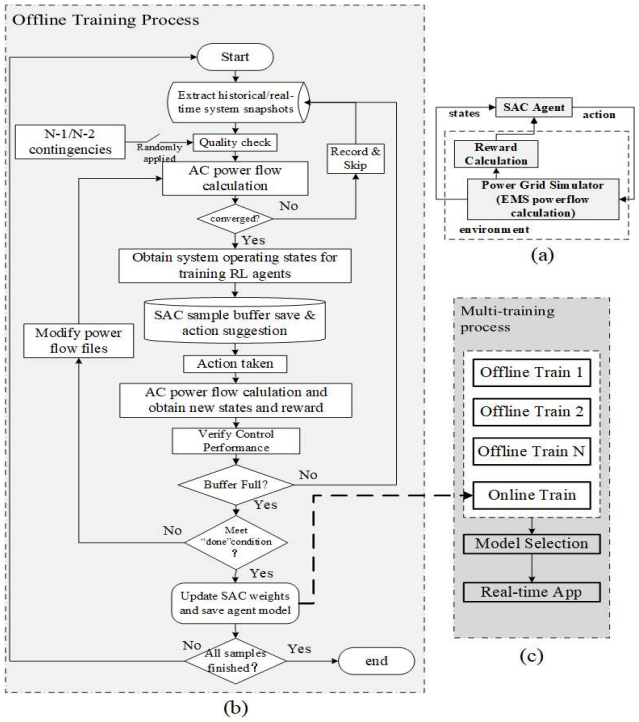
Proposed Solution

Deriving multi-objective real-time control actions is formulated as Markov Decision Process. The soft actor-critic (SAC) algorithm is adopted to solve this control problem because of its superior performance in fast convergence and robustness. The main flowchart of the proposed methodology contains three key modules:

Module (a): provides the interaction process between the power grid simulation environment, an AC power flow solver satisfying Eq. (2) through Eq. (7), and the SAC agent. The environment receives control actions, outputs the corresponding next system states and calculates the reward; while the SAC agent receives states and reward before outputting control actions, in order to satisfy Eq. (8), Eq. (9) and minimize the objective function, Eq. (1).

Module (b): shows the offline training process of an SAC agent. Representative power grid operating snapshots are collected from EMS for preprocessing. System state variables are extracted from those converged snapshots and fed into SAC agent training module, where neural networks are used to establish direct mappings between system states and control actions.

Module (c): To ensure long-term effectiveness and robustness of SAC agent, multiple training processes with different sets of hyperparameters are launched simultaneously, including several offline training processes and one online training process (initialized by the best offline-trained model). The best-performing model is then used for application in real-time environment.



Training Effective SAC Agents

Control Objectives:
- Fix voltage violations, [0.97,1.07] p.u.
- Reduce transmission losses (220kV+) without overloading transmission lines

Episode
- EMS snapshots, every 5 minutes, full-topology model

State Space
- Bus voltage magnitudes (~50 substations)
- Transmission line flows (~100 lines)

Action space
- Voltage setpoints of 12 generators in 5 power plants

Reward definition

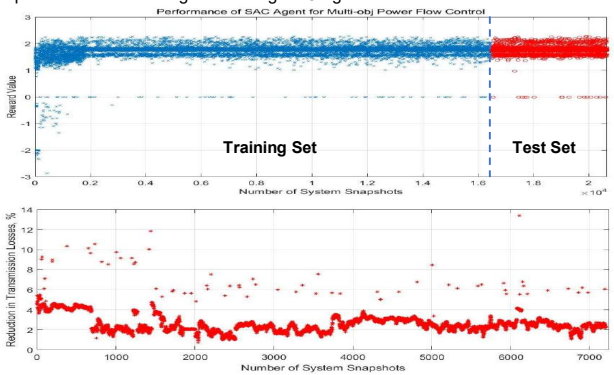
$$reward = \begin{cases} -(\Delta P_{overflow})/10 - (\Delta V_{violation})/100, & \text{if voltage or line flow violation is detected} \\ 50 - \Delta P_{loss} * 1000, & \text{if } \Delta P_{loss} < 0 \\ -100, & \text{if } \Delta P_{loss} \ge 2\% \\ -1 - \Delta P_{loss} * 50, & \text{otherwise} \end{cases}$$

$$\Delta P_{loss} = P_{loss} - P_{loss\_pre} \qquad \Delta P_{overflow} = \sum_{i,j}^{N} ((S_{ij} - S_{ij}^{\max})^2 \qquad \Delta V_{violation} = \sum_{i}^{M} ((V_i - V_i^{\max}) * (V_i - V_i^{\min}))$$

N is the total number of lines with thermal violation; M is the total number of buses with voltage violation; P_{loss} is the present transmission loss value and P_{loss_pre} is the line loss at the base case.

Control Performance of SAC Agents

The presented methodology for multi-objective power flow control was deployed in the control center of SGCC Jiangsu Electric Power Company and tested on a city-level power grid (Zhangjiagang, 220+ kV) with 45 substations, 5 power plants (with 12 generators) and around 100 transmission lines. The performance of training and testing SAC agents is illustrated below:



From 12/3/2019 to 1/13/2020, 7,249 operating snapshots were collected. Three training processes are simultaneously launched and updated twice a week to ensure control performance. For real-time application during this period, the developed method provides valid controls for **99.41**% of these cases. The average line loss reduction is **3.6412%** (compared to the line loss value before control actions). There are 1,019 snapshots with voltage violations, in which SAC agent solves 1,014 snapshots completely and effectively mitigates the remaining 5 snapshots.