CALIFORNIA
ENERGY COMMISSION

Energy Research and Development Division

FINAL PROJECT REPORT

Integrated Distributed Energy Resources Management System

May 2023 | CEC-500-2023-016

PREPARED BY:

Primary Author:
Nanpeng Yu

University of California, Riverside
900 University Ave,
Riverside CA, 92521
(951) 827-3688
https://www.ucr.edu

Contract Number: EPC-15-090

PREPARED FOR:
California Energy Commission

Qing Tian
Project Manager

Reynaldo Gonzalez
Branch Manager
ENERGY SYSTEMS RESEARCH Branch

Jonah Steinbuck, Ph.D.
Director
ENERGY RESEARCH AND DEVELOPMENT DIVISION

Drew Bohan
Executive Director

DISCLAIMER

This report was prepared as the result of work sponsored by the California Energy Commission. It does not necessarily represent the views of the Energy Commission, its employees, or the State of California. The Energy Commission, the State of California, its employees, contractors, and subcontractors make no warranty, express or implied, and assume no legal liability for the information in this report; nor does any party represent that the uses of this information will not infringe upon privately owned rights. This report has not been approved or disapproved by the California Energy Commission, nor has the California Energy Commission passed upon the accuracy or adequacy of the information in this report.

ACKNOWLEDGEMENTS

The authors thank the California Energy Commission for its financial support. The helpful comments and suggestions from technical advisors at the California Energy Commission, Southern California Edison, Pacific Gas and Electric Company, General Electric, and Siemens AG greatly improved the quality of the research.

PREFACE

The California Energy Commission's (CEC) Energy Research and Development Division supports energy research and development programs to spur innovation in energy efficiency, renewable energy and advanced clean generation, energy-related environmental protection, energy transmission and distribution and transportation.

In 2012, the Electric Program Investment Charge (EPIC) was established by the California Public Utilities Commission to fund public investments in research to create and advance new energy solutions, foster regional innovation and bring ideas from the lab to the marketplace. The CEC and the state's three largest investor-owned utilities—Pacific Gas and Electric Company, San Diego Gas & Electric Company and Southern California Edison Company—were selected to administer the EPIC funds and advance novel technologies, tools, and strategies that provide benefits to their electric ratepayers.

The is committed to ensuring public participation in its research and development programs that promote greater reliability, lower costs, and increase safety for the California electric ratepayer and include:

- Providing societal benefits.
- Reducing greenhouse gas emission in the electricity sector at the lowest possible cost.
- Supporting California's loading order to meet energy needs first with energy efficiency and demand response, next with renewable energy (distributed generation and utility scale), and finally with clean, conventional electricity supply.
- Supporting low-emission vehicles and transportation.
- Providing economic development.
- Using ratepayer funds efficiently.

Integrated Distributed Energy Resources Management System is the final report for the Integrated Distributed Energy Resources Management System (iDERMS) project (Contract Number EPC-15-090) conducted by The Regents of the University of California (UC Riverside). The information from this project contributes to the Energy Research and Development Division's EPIC Program.

For more information about the Energy Research and Development Division, please visit the Energy Commission's research website (www.energy.ca.gov/research/) or contact the CEC at 916-327-1551.

ABSTRACT

This project developed an Integrated Distributed Energy Management System that coordinates a large number of individual distributed energy resources in utility electricity-distribution systems. This Integrated Distributed Energy Management System report includes three key software modules: three-phase optimal power flow, Volt-VAR control, and network reconfiguration. Scalable and efficient three-phase optimal power flow modules essentially determined the best-available generation and load dispatch in each of the distribution feeders. The data-driven Volt-VAR control module adjusted voltage-regulating device settings to both reduce network losses and maintain customer voltage within reasonable bounds. The network-reconfiguration algorithm also reduced both network losses and outage durations. The performance of all three software modules was successfully validated through distribution test feeders using real-world smart-meter data. Deliverables from this project can be easily adopted by energy-industry software vendors and incorporated into advanced distribution-management systems. Once this technology is implemented in electricity distribution systems, California's utility ratepayers will benefit from lower electricity bills, higher voltage quality, and greater system reliability.

Keywords: Distributed Energy Resources, Advanced Distribution Management System, Data-Driven Control

Please use the following citation for this report:

Yu, Nanpeng. 2020. *Integrated Distributed Energy Resources Management System (iDERMS)*. California Energy Commission. Publication Number: CEC-500-2023-016.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Introduction

Senate Bill (SB) 100, signed into law by Governor Jerry Brown in 2018, mandates that renewable energy and zero-carbon resources supply 100 percent of electricity retail sales to end-use customers by the end of 2045. California is already experiencing significant and growing penetration of distributed energy resources. For example, a distribution feeder serving a few thousand customers can have thousands of distributed energy resources such as rooftop solar photovoltaic (PV) units, wind turbines, fuel cells, battery storage, electric vehicles, and EV chargers. It is, therefore, challenging to manage operations of large numbers of independent distributed energy resources in a coordinated manner. Current distribution management systems are incapable of dispatching thousands of distributed energy resources while maintaining customer voltage and minimizing electricity costs, outage frequencies, and their durations. This project developed an integrated distributed energy resources management system capable of managing higher penetration of the state's distributed renewable resources required to achieve California's ambitious renewable energy resource goals.

There are two primary technical barriers to the successful development of distributed energy resources management systems. The first is the scalability of existing distribution systems: distributed energy resource modeling, monitoring, and control algorithms. The computation time of conventional algorithms does not scale well with either the number of system connection points (nodes) or the number of distributed energy resources. The second barrier is the lack of complete and accurate utility distribution system data. A regional utility's distribution system contains millions of connection points. It is extremely time consuming and labor intensive for electric utilities to establish an entirely accurate database that documents the connectivity between the customers and service transformers.

Project Purpose

The main objective of this project is to develop an Integrated Distributed Energy Management System capable of expanding and managing high penetration of renewable resource generation. The project team sought to advance distributed energy resource integration into the state's electric distribution systems through both the project's large-scale decentralized applications and control algorithms within a data-driven control framework. This innovative project technology development, if broadly adopted, could transform the way utility customers and distributed energy resources interact with their electricity distribution systems. Newly decentralized distribution system controls and a proactive resource-participation scheme could ultimately replace traditional centralized control and current passive customer-demand-response.

This project is the first distribution automation system to apply a decentralized control concept that coordinates distributed energy resource operations to reduce costs and enhance grid reliability and resiliency. The three-fold project goals are to:

- Address the challenge of consistently coordinating management of large numbers of diverse and independent distributed energy resources.
- Increase renewable-resource penetration by mitigating generation uncertainties with innovative methods including decentralized Volt-VAR control and three-phase optimal power flow.
- Improve transmission and distribution grid reliability with advanced electric distribution system reconfigurations and restoration algorithms.

The proposed Integrated Distributed Energy Management System overcomes the technical barriers to integrating more renewable resources generation into utilities' existing electricity systems. Its technical solutions are extremely scalable. By developing a decentralized algorithm, researchers significantly improved the efficiency of existing distributed energy resource dispatch algorithms. By taking a data-driven approach to controlling the distribution system with distributed energy resources, the technology's implementation is simplified since it does not require complete and accurate distribution system models. If implemented, the Integrated Distributed Energy Management System RMS could coordinate distributed energy resource operations and increase local renewable resource penetration in California utilities' respective service territories. Power-system software vendors like General Electric Company and Siemens AG could integrate the three key software modules into their commercial advanced distribution management system products.

Project Approach

The project team first developed a theoretical model of core algorithms for the three Integrated Distributed Energy Management System MS modules. Team researchers then developed software simulations to verify proposed algorithms through a combination of standard power distribution test feeders and real-world smart-meter data. Team members specifically used a comprehensive list of distribution-system test feeders, including radial and meshed topology and various connection points. Performance of the proposed algorithms was rigorously measured against benchmark algorithms, which the power engineering industry considers to be state-of-the-art. Lastly, researchers integrated the three modules, the three-phase optimal power flow, the Volt-VAR control, and network reconfiguration into the Integrated Distributed Energy Management System platform.

The project team encountered two technical barriers during the project. The first was how to develop an efficient and workable solution for the three-phase optimal power flow problem for large-scale distribution circuits. To overcome this barrier, researchers combined an advanced algorithm with a decentralized technique to solve the non-convex optimization problem. The second technical challenge was how to develop a safe and data-driven algorithm for Volt-VAR control and network reconfiguration. The researchers countered this technical challenge by rigorously modeling the operational constraints of the electric distribution system. A technical advisory committee for this research project included representatives from Southern California Edison, Siemens AG, General Electric, and the California Independent System Operator. The technical advisory committee provided feedback and constructive insights, further suggesting that the best path for dissemination of the technology and its ultimate commercialization would be through close collaboration with software vendors like General Electric and Siemens AG since those vendors already develop advanced distribution-management and distributed-energy resources management systems for electric utilities.

Project Results

All three project goals were met. The project team successfully developed the Integrated Distributed Energy Management System, which coordinates operations of large-scale distributed-energy resources. The performance of the Integrated Distributed Energy Management System was successfully validated through simulations on the Institute of Electrical and Electronics Engineers' standard test feeders with real-world, synthetic smart-meter data. The three-phase optimal power-flow module significantly improved the scalability and reduced the time required for existing work. The simulation results also showed that the proposed Volt-VAR control module can reduce loss reductions while maintaining voltages of all connection points within reasonable bounds. The proposed Volt-VAR control algorithm outperformed existing algorithms by eliminating reliance on complete and accurate distribution system topology. The numerical study results demonstrated that the proposed distribution system reconfiguration algorithm can produce a more effective and efficient system configuration, which also reduces network losses. This technology filled essential knowledge gaps needed to significantly increase penetration of distributed energy resources into electricity distribution systems. The analysis also identified the need for electric utilities to carefully collect and analyze customer electricity consumption and voltage data from utility smart meters.

The software modules and algorithms developed in this project are available on GitHub (a software repository hosting site) and the project website. The technologies developed in this project outperformed all existing algorithms reported in the literature.

A major lesson learned is that electric utilities typically do not have reliable distribution network topology and parameter data. This restriction limits applicability of model-based distribution-system control algorithms. The Integrated Distributed Energy Management System platform alleviates this issue by using data-driven distribution system control algorithms that do not depend upon accurate distribution network topology and parameter data.

More research is required to further knowledge levels and advance the technology's maturity. The Integrated Distributed Energy Management System requires that all smart meters transfer their data to the distribution system control center in real time. However, some electric utilities may have limited communication bandwidth for their advanced-metering infrastructures. There is, therefore, a need to develop multi-agent data-driven distribution-system control algorithms that do not rely upon a centralized distribution system.

Technology/Knowledge Transfer/Market Adoption (Advancing the Research to Market)

The project team shared knowledge gained, and technologies developed in this project in three ways. First, the technologies developed in this project are summarized in journals and conference papers [16-18] that are disseminated to academia. Second, project team members presented the research results in conferences for electric utilities. Third, the software modules developed in this project are available on the project's website (https://intra.ece.ucr.edu/~nyu/papers/SoftwareCode).

Energy industry software vendors like Siemens AG and General Electric have shown interest in adopting data-driven control algorithms similar to those developed in this project. The target market for this technology is electric utilities. The project team anticipates that all electric utilities in California will eventually adopt data-driven distribution network control algorithms.

The reviewers of the conference and journal papers [16-18] that featured this project's results all provided very positive feedback that praised the scalability of the proposed approaches. Southern California Edison strongly recommended that the project team work collaboratively with General Electric to implement the proposed algorithms for advanced distribution-management systems. The project team plans to continue commercialization discussions with energy-industry software vendors like General Electric and Siemens AG.

Benefits to California

The technology developed through this research can significantly increase penetration of distributed energy resources in California, which could reduce greenhouse gas emissions and customer utility bills. The Integrated Distributed Energy Management System can significantly enhance grid reliability, reduce electricity costs, and improve safety. Specifically, the three-phase power flow which has the objective of minimizing operational cost, improves energy dispatch efficiency, which in turn can reduce customer electricity costs. It also ensures that the distribution system operates within appropriate voltage limits. The data-driven Volt-VAR control algorithm reduces peak feeder loads and prevents voltage irregularities. The distribution network reconfiguration and restoration technology together enhance distribution system reliability by anticipating the unfavorable renewable and load dynamics that could cause system disturbances and potential outages.

This research shows that the proposed three-phase optimal power flow module of the Integrated Distributed Energy Management System platform could reduce energy-dispatch costs by up to 10 percent. The proposed data-driven Volt-VAR control algorithm could reduce distribution-system losses and operational costs of voltage-regulating devices by 10 percent.

If all electric utilities in California, theoretically, adopted this proposed technology, potential annual savings could top $4.28 billion. The technology could also be adopted by microgrid operators to coordinate their distributed energy resource operations and maintain service voltages. Generation of up to 28,549 gigawatt-hours of electricity could be avoided due to reduced network losses. Potential reductions of greenhouse gases could be as great as 13,103,991 metric tons, which is calculated based on the standardized emission factors for electricity of 0.000283 metric tons/kWh.

This research has set the groundwork for more extensive capacity analyses on electric-distribution systems in California.

CHAPTER 1: Introduction

This project focuses on three concepts: electric distribution-system optimal power flow, Volt-VAR control, and network reconfiguration.

Distribution System Optimal Power Flow

Optimal power flow (OPF) is within a class of optimization problems in electric power system engineering. OPF seeks to streamline operations of electric power systems subject to physical constraints imposed by natural electrical laws and engineering limits. For example, a classic OPF problem asks: What is the optimal way (in the sense of minimum generation costs) to schedule the output of a set of power generators so that all loads are served, and no voltage or current operating limits are violated? The OPF problem can be solved as a mathematical programming problem with mathematical programming algorithms. The same OPF program is also capable of computing system marginal costs. For example, generation costs stemming from bus active-power injection changes can be used as nodal prices for pricing transmission services since they reflect both transmission loss and congestion components for transferring electricity from one point to another [2]. This enables a market approach for transmission-congestion management.

As distributed energy resources (DERs) and smart buildings increasingly penetrate existing electric-power distribution systems, the dynamic resource management of thousands of DERs becomes difficult. This difficulty can be overcome with a distribution system market approach where electric utility customers proactively participate in resource dispatch and price formation processes achieved through incorporation of the OPF concept into electric distribution systems. This market approach is supported by many researchers and policy makers. For example, the New York Public Service Commission kicked off a proposal called Reforming the Energy Vision (REV), which trains distribution-system operators (DSOs) who coordinate and implement planning and operations for DERs and smart buildings [3].

The mathematical formulation of the OPF problem is a complex, nonlinear programming problem. Solution algorithms include: Newton’s method, linear and quadratic programming, nonlinear and polynomial programming, interior point methods, semi-definite programming, and heuristic optimization methods. Theoretical guarantees of solution optimality, however, are only available to single-phase and tree networks. Electric distribution systems are unbalanced because of unbalanced loads across three phases and untransposed feeder lines. This forces distribution-system modeling to be three-phase rather than single-phase, exacerbating an already challenging problem.

One of the main goals of this project was to develop an efficient and scalable three-phase OPF algorithm for electric distribution systems capable of identifying global solutions. Researchers first revisited the conundrum of solving three-phase OPF problems. A counter example of a three-phase network showed that a solution could not be guaranteed with the semi-definite programming (SDP) relaxation method. To efficiently find a global solution, this project proposed an innovative three-phase OPF algorithm by combining the convex-iteration technique with the chordal-based conversion algorithm. The researchers also proposed an algorithm to develop a grid-partitioning scheme to reduce computational complexity. Numerical simulations were conducted on the IEEE test feeders to validate the computational efficiency and scalability of the proposed algorithm and solutions. The simulation results showed that the proposed algorithm can be globally feasible even when the SDP relaxation method fails. The partition algorithm effectively identified a chordal conversion that made the overall algorithm computationally efficient. Finally, the simulation results from the IEEE 123-bus and 906-bus test feeders demonstrated the scalability of the proposed algorithm.

Volt-VAR Control

One of the primary goals of a distribution management system is to maintain system-wide voltage levels and reactive power flows (though voltage levels can vary by a small amount from nominal values depending on the electric resistance and reactance of power-system devices as well as current-flow magnitudes). For example, nominal system voltages for U.S. residential customers are 120 (phase-to-neural) and 240 volts (phase-to-phase). During normal operation, typical Southern California utility customer's daily voltages can vary between 225 and 252 volts. Large-voltage deviations from nominal values can damage utility equipment and customer loads, so system voltages must be regulated. For example, the American National Standards Institute (ANSI) C84.1 voltage ranges require that utilities design electric systems that provide service voltages within± 5 percent of the nominal values with infrequent excursions, and-8.3 percent to $+5.8$ percent with limited frequency and duration [4].

In recent years, increasing numbers of DERs have been added to medium- and low-voltage-level distribution feeders. Due to the uncertain power output of this renewable generation, regulating voltages can be problematic. Grid-connected DERs function as distributed generation, which boosts voltage levels at nearby locations. Conventional control methods therefore face inconsistent control objectives. For instance, when some locations have high DER penetration and others do not, raising voltage levels in low-voltage regions will cause overly high voltage in originally high-voltage regions, and vice versa.

To tackle the issue of managing distribution system-wide voltage levels, Volt-VAR control (VVC) was developed and integrated into the distribution management system. VVC determines the best set of control actions for all voltage regulations and VAR control devices (that is, voltage regulators, on-load tap changers, and switchable capacitor banks), to reduce system losses and equipment operating costs without violating operating constraints such as voltage limits.

Existing VVC algorithms mainly adopt a physical model-based control approach. However, the VVC problem is often solved using mathematical programming or trial-and-error methods. Physical model-based approaches rely on accurate knowledge of distribution grid topologies and parameters like line impedances. However, it is difficult for regional electric utilities to maintain reliable network models, which often involve millions of buses in primary and secondary feeders. To cope with incomplete models, VVC actions could be tried out to determine the greatest reward. This project developed a novel deep-reinforcement learning (DRL) algorithm named Constrained Soft Actor-Critic (CSAC) that enables data-driven and model-free implementation of VVC. This algorithm determines a near-optimal control policy of devices from operational data without relying on complete and accurate distribution network topology and parameter information. In contrast to existing DRL-based methods, this algorithm determines a control policy that directly selects control actions instead of consulting an action-value function. This is particularly useful for VVC problems since it is much simpler to approximate control-policy functions than action-value functions. The proposed CSAC algorithm also explicitly models physical operation constraints by combining the merits of multipliers and a soft actor-critic (SAC) [5] algorithm; the proposed CSAC algorithm can better satisfy operation constraints in power-distribution systems. Finally, the algorithm is off-policy, meaning it is more sample-efficient than state-of-the-art DRL algorithms for constrained Markov decision process problems. By using an ordinal network structure to encode the natural ordering between discrete actions of voltage-regulating devices and introducing a device-decoupled policy network structure, this algorithm demonstrates significant improvements over existing DRL-based methods for sample efficiency and scalability.

Distribution Network Reconfiguration

Distribution network reconfiguration refers to changing a network's topology by changing the status of remotely controllable switches (RCSs). Such distribution automation is traditionally used for service restoration [6] where a portion of the distribution network is affected by events such as faults or scheduled maintenance. The network reconfiguration isolates the affected region by opening the surrounding switches so that the corresponding line segments no longer carry any current. After the fault is cleared or the maintenance completed, some switches can be closed to re-connect the isolated portions to adjacent feeders. Alternatively, distribution network reconfiguration (DNR) can also be used to improve service operational criteria other than service restoration: for example, loss minimization. In recent years, a growing number of DERs have been installed in medium- and low-voltage distribution feeders. High penetration of DERs can cause reversed power flow and change the network I^2R loss pattern. Distribution network reconfiguration, in this case, is one of the most effective operational strategies for loss minimization.

DNR can be performed either statically or dynamically. The former concerns determining and fixing the best configurations for the entire study period; the latter determines a sequence of hourly configurations over time. The mathematical formulations of DNR problems are typically mixed-integer-programming (MIP) models, where integer variables represent switch status. Formulations for the dynamic DNR are further characterized by three elements. First, the number of switching actions needs to be constrained to reduce device wear and tear. Second, the problem size is typically much larger than the static DNR since multiple time steps need to be worked out simultaneously. Third, dynamic DNR requires the modeling of load uncertainties, for instance by point estimations or uncertainty sets.

Most of the existing literature on this dynamic DNR problem adopts a physical model-based control approach and focuses on solving a mixed-integer program. There remain, however, several technical limitations with this approach. First, physical model-based formulations can be difficult to adopt in practice due to model uncertainty. Uncertainties in real-world systems include not only DER loads but accurate network models. In particular, primary and secondary networks' parameter estimates are difficult to maintain by electric utilities. Second, for mixed-integer programming techniques and meta-heuristics algorithms the computation time increases substantially with the size of the network and the length of the planning horizon; so the algorithm may fail to converge within a reasonable time.

To cope with unreliable distribution network parameters and the long computation time, project researchers propose a deep reinforcement learning (RL) framework to learn and execute dynamic DNR without using distribution network parameter information. One of the major limitations of existing RL algorithms is the low sample efficiency. To address this, researchers propose an innovative approach to augment past grid operational experiences with synthetic ones. The proposed off-policy RL algorithm is capable of performing dynamic DNR with only network topology information and an historical operation data set. This operation experience augmentation technique improves the performance of the RL algorithm.

CHAPTER 2: Project Approach

iDERMs Platform

The Integrated Distributed Energy Management System (iDERMs) is the intelligent distributed energy resources management system that enhances distribution network control and coordinates distributed energy resources. It improves the optimality, stability, and reliability of distribution system management. The software package was developed in MATLAB and Python.

Functionality

The software package contains three main modules including the three-phase optimal power flow, the network reconfiguration, and the Volt-VAR distribution-system control.

Three-Phase OPF

As DERs and smart buildings increasingly penetrate electric distribution systems, dynamic resource management on large-scale systems with thousands of DERs becomes difficult. This difficulty can be countered with a distribution system market approach where electricity customers can proactively participate in the resource dispatch and price formation processes. The operation of distribution-system markets relies on the three-phase OPF algorithm. This algorithm is used to identify the most efficient DER dispatch that also minimizes total generation cost while satisfying operational constraints, including voltage and thermal-limit constraints. The proposed problem is a rank-constrained SDP problem and was solved with the convex iteration algorithm with chordal conversion.

Network Reconfiguration

The dynamic DNR performs hourly dynamic status changes of sectionalizing and tie switches to reduce network line losses, minimize load loss, or increase hosting capacity for distributed energy resource generation. Deep Q-learning with data augmentation addresses the distribution network reconfiguration problem. The algorithm has three components: deep Q-learning, radial-configuration discovery, and operational experience augmentation.

Volt-VAR Control

VVC plays an important role in enhancing energy efficiency, power quality, and reliability of electric distribution systems by coordinating the operations of equipment such as voltage regulators, on-load tap changers, and capacitor banks. VVC both maintains voltage in the distribution system within desired ranges and reduces system operation costs, which include network losses and equipment depreciation from wear and tear. The data-driven algorithms are adopted to tackle the VVC problem. The VVC is proposed as a constrained Markov decision process (CMDP) and solved with the policy-gradient-based algorithms.

Graphic User Interface

The graphic iDERMS interface platform was developed in Python. For each module of the three main functionalities, different algorithm and test cases can be selected from the scroll menu. The environment for MATLAB or Python can be set through the text input box. The example panels for the three modules are shown in Figure 1 through Figure 3.

Three-Phase OPF

The algorithm can be selected from the proposed convex iteration algorithm, sequential quadratic programming (SQP), and interior-point method (IPOPT). The test feeders contain the IEEE bus-4, bus-10, bus-13, bus-34, bus-37, bus-123, and bus-906 distribution networks.

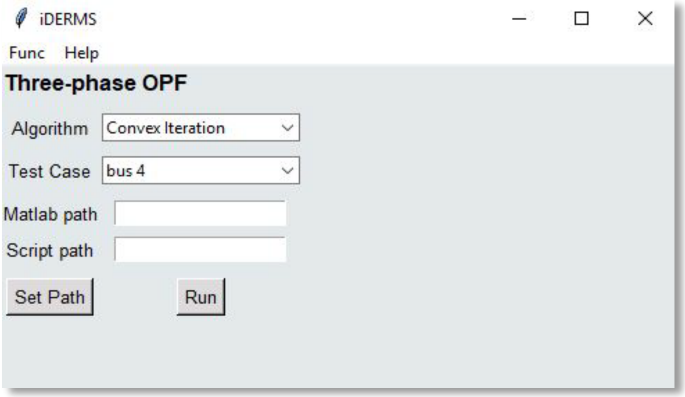
Network Reconfiguration

The available algorithm includes the DQN and the proposed augmented DQN method. The validation is performed on the bus-16 test feeder.

Volt-VAR Control

The algorithms, including the proposed constrained SAC, CPO, and DQN, can be selected. The experiments can be conducted on the IEEE bus-4, bus-34, and bus-123 distribution feeders.

Figure 1: Graphic Interface Panel for OPF Module



Three-phase OPF

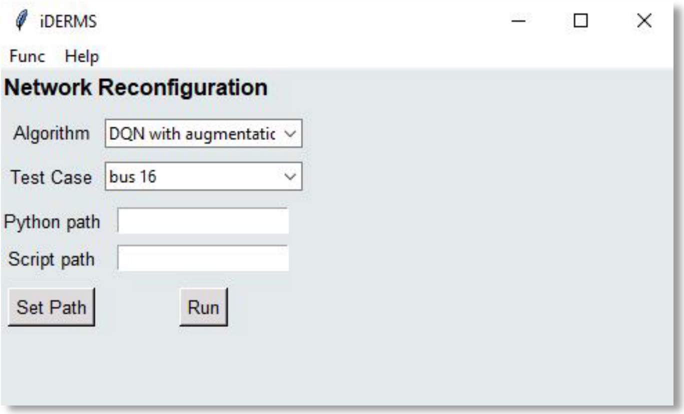
Algorithm: Convex Iteration

Test Case: bus 4

Matlab path:

Script path:

Figure 2: Graphic Interface Panel for Network Reconfiguration Module



Network Reconfiguration

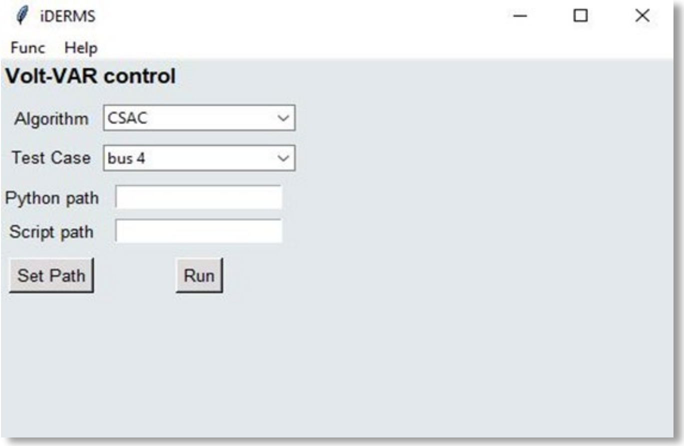
Algorithm: DQN with augmentatic

Test Case: bus 16

Python path

Script path

Figure 3: Graphic Interface Panel for Volt-VAR Control Module



Volt-VAR control

Algorithm: CSAC

Test Case: bus 4

Python path

Script path

Source: Yuanqi Gao, Jie Shi, Wei Wang and Nanpeng Yu, "Dynamic Distribution Network Reconfiguration Using Reinforcement Learning," IEEE SmartGridComm, pp. 1-7, 2019.

Three-Phase Optimal Power Flow

Objective

The volume and diversity of DERs are growing rapidly, increasing the need to ease their penetration into the state's electric power distribution system. The distribution system market is addressing the difficulties of coordinating these DERs, so that utility customers can proactively participate in the resource dispatch and price formation processes. To establish this distribution system market, the three-phase OPF problem needed to be solved. In this project, an efficient and scalable three-phase OPF algorithm was developed.

Framework

The overall framework of the three-phase OPF algorithm is shown in Figure 4. The framework consists of three main elements:

- Rank-constrained SDP relaxation.
- Grid partition.
- Convex iteration.

Figure 4: Overall Framework of Three-Phase OPF Algorithm

Rank-constrained SDP Relaxation
Grid Partition
Convex Iteration

Source: Yuanqi Gao, Jie Shi, Wei Wang and Nanpeng Yu, "Dynamic Distribution Network Reconfiguration Using Reinforcement Learning," IEEE SmartGridComm, pp. 1-7, 2019.

Rank-Constrained SDP Relaxation

The three-phase OPF problem is basically a non-linear, non-convex optimization problem. The linearization technique adopted for a single-phase OPF problem in an electric transmission system is no longer sustainable because of distribution line resistance and the unbalanced characteristics of a three-phase load. This semi-definite programming approach's tight relaxation is attracting commercial interest. The three-phase OPF problem can be reformulated into a rank-constrained SDP problem:

$$\min_{\mathrm{X}} C(X)$$

$$\text{s.t. } \mathrm{X} \in \mathrm{B}$$
$$\mathrm{X} \succcurlyeq 0$$

$$\mathrm{rank(X)} = 1$$

$\mathrm{X} = \mathrm{VV^T}$, and V is the vector of nodal voltage variables. B is the feasible region of X. The last rank constraint is the only non-convex constraint. The existing relaxation approaches typically directly drop the rank constraint, which could result in a non-feasible solution. A bi-linear penalty term is introduced in the approach in the objective function to further tighten the relaxation.

$$\min_{\mathrm{X,W}} C(X) + \lambda Tr(XW)$$

λ is the penalty coefficient. W is the direction matrix pointing to the rank-one sub-space, and $\mathrm{I} \succcurlyeq \mathrm{W} \succcurlyeq 0$. If the rank-one solution of X is achieved, the penalty term will become zero. For a traditional bilinear optimization problem, an iterative linear programming method can be applied to find the optimal solution(s). In the context of SDP, the convex iteration algorithm is proposed as:

Step 1:

$$\min_{X} c(x) + \lambda Tr(XW)$$

$$\text{s.t. } X \in \text{B}$$

$$X \succcurlyeq 0$$

Step 2:

$$\min_{W} Tr(XW)$$

$$\text{s.t. I} \succcurlyeq W_l \succcurlyeq 0$$

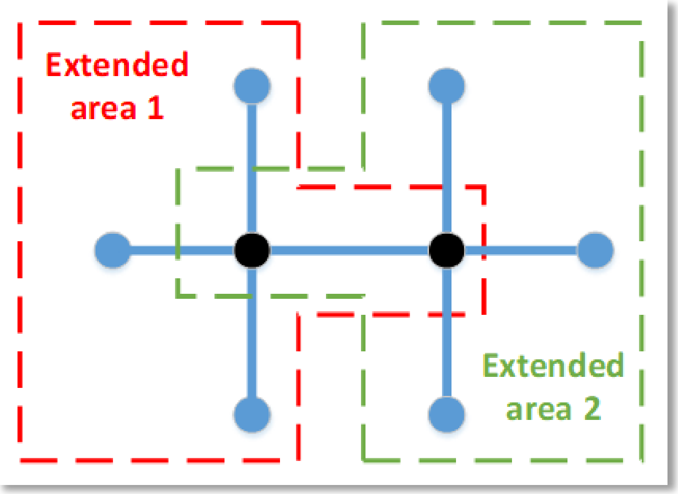
$$\text{Tr}(W) = N_X - 1$$

$\mathrm{N_X}$ is the dimension of X, and the X and W are solved iteratively by using the result of W or X from the previous step. The direction matrix W can be initialized as an identity matrix.

Grid Partition and Chordal Conversion

By directly formulating the three-phase OPF problem into an SDP problem, the number of decision variables is $\mathrm{N}^2$where N is the number of nodes in the distribution system. Therefore, the complexity of solving the problem will increase rapidly with the size of the network. However, the distribution system is typically a tree network. The sparsity of the voltage variable matrix can be exploited with the chordal conversion.

Figure 5: Example of Grid Partition



Source: Yuanqi Gao, Jie Shi, Wei Wang and Nanpeng Yu, "Dynamic Distribution Network Reconfiguration Using Reinforcement Learning," IEEE SmartGridComm, pp. 1-7, 2019.

The semi-definite completion theorem states that a symmetric matrix is positive and semi-definite completable if and only if all of the small matrices associated with the maximal cliques of the graph derived from the whole matrix are positive and semi-definite. This property allows the SDP problem to be converted into another form with smaller-sized positive semi-definite variables by portioning the whole network into extended areas. In this way, the necessary variables can be significantly reduced.

$$\min_{\mathrm{X}} \sum_{l=1}^{N_A} C_l(X_l^{ext})$$

$$\begin{aligned} \mathrm{s.t.}\ \mathrm{X_l^{ext}} &\in \mathrm{B}^{(l)}, \mathrm{l} = 1{,}2..\mathrm{N_A} \\ \mathrm{X_l^{ext}} &\succcurlyeq 0, \mathrm{l} = 1{,}2..\mathrm{N_A} \\ \mathrm{X_l^{ext(r)}} &= X_r^{ext(l)}, \mathrm{l}, \mathrm{r} = 1{,}2..\mathrm{N_A} \\ \mathrm{rank}(\mathrm{X_l^{ext}}) &= 1, \mathrm{l} = 1{,}2..\mathrm{N_A} \end{aligned}$$

Extra auxiliary variables have to be introduced for the boundary of the connected sub-networks to enforce the voltage equality of the shared nodes. $\mathrm{X_l^{ext}}$ is the variable associated with the l-th extended area. $\mathrm{X_l^{ext(r)}}$ is the variable associated with area of the l-th extended area that intersects the r-th extended area. An example of grid partition is shown in Figure 5. The whole network is divided into to two sub-areas, where the black dots are the shared nodes.

Chordal Conversion Based Convex Iteration

By synergistically combining the chordal conversion method and the convex iteration technique, a new iterative three-phase OPF solution algorithm is proposed:

Step 1:

$$\min_{\mathrm{X}} \sum_{l=1}^{N_A} C(X_l^{ext}) + \lambda_i Tr(X_l^{ext}\, W_l)$$

$$\begin{aligned} \mathrm{s.t.}\ \mathrm{X_l^{ext}} &\in \mathrm{B}^{(l)}, \mathrm{l} = 1{,}2..\mathrm{N_A} \\ \mathrm{X_l^{ext}} &\succcurlyeq 0, \mathrm{l} = 1{,}2..\mathrm{N_A} \\ \mathrm{X_l^{ext(r)}} &= X_r^{ext(l)}, \mathrm{l}, \mathrm{r} = 1{,}2..\mathrm{N_A} \end{aligned}$$

Step 2:

For each extended area $, \mathrm{l} = 1{,}2..\mathrm{N_A}$:

$$\min_{\mathrm{W}} Tr(X_l^{ext}\, W_l)$$

$$\begin{aligned} \mathrm{s.t.}\ \mathrm{I} \succcurlyeq W_l &\succcurlyeq 0 \\ \mathrm{Tr}(\mathrm{W_l}) &= N_{X_l}^{ext} - 1 \end{aligned}$$

where X_l^{ext} is with size $N_{X_l}^{ext} \times N_{X_l}^{ext}$. $\mathrm{W_l}$ is the direction matrix for l-th extended area, which has a closed-form solution for the problem in step 2:

$$W_l = U_j\left(:,2:N_{X_l^{ext}}\right) U_j\left(:,2:N_{X_l^{ext}}\right)^T$$

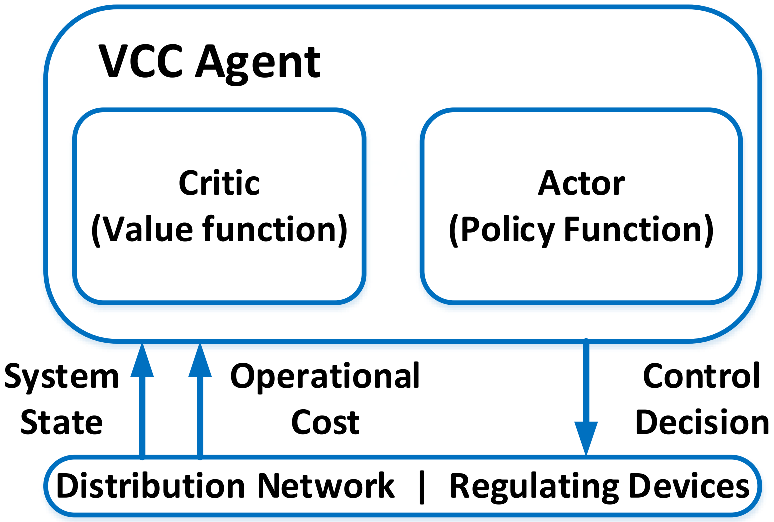
U_j is obtained from singular-value decomposing $\mathrm{X_l^{ext}} = U_j \Lambda_j U_j^T$. The algorithm repeatedly solves X by fixing W, then solves W until the trace penalty terms converge to zero, which means a rank-one global optimal solution X is achieved.

Data Driven Volt-VAR Control

Objective

Voltage profiles highly impact electricity service quality for utility end users. Both over-voltage and under-voltage conditions could reduce energy efficiency, cause equipment malfunctions, and damage customers' electrical appliances. Equipped with remote control and monitoring devices, electric utilities started adopting VVC to maintain voltages within an allowable range, manage the power factor, and reduce operation costs. These control objectives can be achieved by coordinating the operation of various equipment such as voltage regulators, on-load tap changers, switchable capacitor banks, and smart inverters. However, the lack of robust distribution network topology and parameter information impede wide deployment of existing optimization-based VVC approaches. In this project, a model-free deep reinforcement learning-based VCC algorithm is proposed.

Figure 6: Overall Framework of Deep Reinforcement Learning Based VCC



Source: *Yuanqi Gao, Jie Shi, Wei Wang and Nanpeng Yu, "Dynamic Distribution Network Reconfiguration Using Reinforcement Learning," IEEE SmartGridComm, pp. 1-7, 2019.*

Framework

The overall framework of the deep reinforcement learning-based VCC is shown in Figure 6. This VCC agent and the distribution grid interact at each of a sequence of discrete time steps $\mathrm{t} = 0,1,2 \ldots.$ At each time step, the agent receives the system's state $\mathrm{s_t}$, and selects a control action $\mathrm{a_t}$. One time step later, the agent receives the reward $\mathrm{R}(s_t, a_t, s_{t+1})$ and the operational cost $R_C(s_t, a_t, s_{t+1})$, and becomes a new state. The goal of the agent is to learn a control policy for the VCC problem.

CMDP Algorithm

The VCC problem is formulated as a CMDP. The VCC agent attempts to learn a policy that maximizes its expected discounted return while restraining its expected discounted cost within the limited budget.

$$\max_{\pi} J(\pi) = \mathrm{E}_{\tau \sim \pi}[\sum_{0}^{T} \gamma^t R(s_t, a_t, s_{t+1})]$$

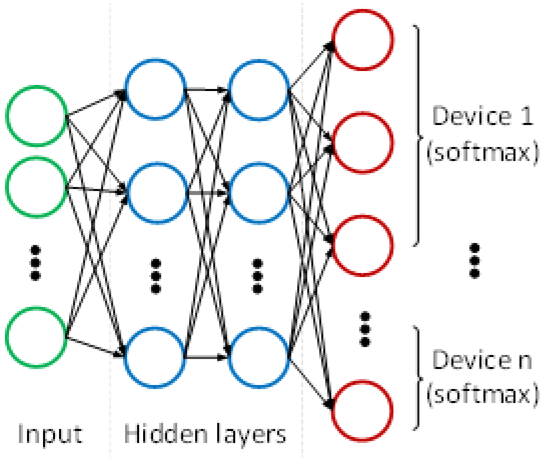
$$\mathrm{s.t.\,J_C}(\pi) = \mathrm{E}_{\tau \sim \pi}[\sum_{0}^{T} \gamma^t R(s_t, a_t, s_{t+1})] \leq \bar{J}$$

The reward R is defined as the negative operational cost, including the costs of system losses and device switching. The cost $\mathrm{R_C}$ is defined in terms of the number of voltage violations across all the network nodes. The constrained policy optimization algorithm is adopted, which statistically guarantees every control policy during learning will satisfy operational constraints in the form of expectation.

Device-Decoupled Policy Function

In a VVC problem, the network loss is determined by the tap positions of all controllable devices together. The number of feasible control actions increases exponentially with the number of controllable devices. However, the control action of regulating devices can be taken independently. Compared with the action value method with epsilon-greedy action choices, the policy gradient method, which directly learns the parameterized policy, could be more scalable with the device-decouple policy network proposed in Figure 7.

Figure 7: Device-Decoupled Policy Network



Source: Wei Wang, Nanpeng Yu, Jie Shi and Yuanqi Gao, "Volt-VAR Control in Power Distribution Systems with Deep Reinforcement Learning," IEEE SmartGridComm, pp. 1-7, 2019.

The output layer of the policy network is with size $\sum_{\mathrm{i}=1}^{\mathrm{N}} n_i$, where N is the total number of regulating devices, and $\mathrm{n_i}$ is the number of tap positions of device i. The probabilities π_i of i-th device to select different tap positions are obtained by applying a soft-max activation function to the i-th subset of the neurons with size $\mathrm{n_i}$. The final probability for the action combination of all the devices is $\pi = \prod_{i=1}^{N} \pi_i$.

Reinforcement Learning-Based Distribution Network Reconfiguration

Objective

Distribution network reconfiguration (DNR) [7] is recognized as one of the most effective methods for improving the distribution system's operational performance under increasing penetration by DER generation. Such performance improvements include DER hosting capacity, minimizing distributed generation curtailments, and minimizing network resistive losses. DNR changes the open/closed statuses of remotely controllable switches (RCS) on the primary or secondary feeder to alter the topology (configuration) of the distribution network for improving a single or multiple operational objectives such as loss minimization. To determine which RCS to open or close to achieve the most effective topology, one could solve the corresponding mathematical programming problem that models the DNR, using mathematical optimization algorithms or other trial-or-error methods. However, implementing these methods requires understanding of network parameters (models) such as resistance and reactance of each line segment of the network, which can be difficult for regional electric utilities to maintain. In this report, researchers describe a reinforcement learning framework that performs network reconfigurations without model-based calculations; instead, the framework provides effective reconfiguration of the network from historical operational data from electric-utility databases. The next subsection formally states the distribution-network reconfiguration (DNR) problem. Following are the subsections that describe the reinforcement-learning-based DNR.

Problem Statement

This subsection describes details of the dynamic distribution network reconfiguration (DDNR) problem. The goal of the dynamic distribution network reconfiguration is to minimize network total-resistive loss (I^2R loss) by changing the status of remotely controllable switches (RCS). Also, to prevent wear-and-tear of the devices, the number of open/closed control actuation of the RCS is regulated as well. As a result, the goal is to minimize:

$$\sum_{l} r_l I_l^2 + w|\alpha_l - a_{l0}| \tag{1}$$

Where r_l is the resistance of line segment l; I_l^2 is the electric current magnitude squared for line segment l. $\alpha_l = 0 \; or \; 1$ is a binary variable representing the open/closed status of the RCS on line segment l (one switch corresponds to one line segment). The term α_{l0} is the existing status of switch l. w is a constant parameter that controls how the devices' wear-and-tear are valued over network loss minimization.

(1) corresponds to the single-time step-loss minimization. In practice, the current flowing on each line segment changes from time to time and the reconfiguration should be with respect to a period of time. The researchers consider minimization of loss over the time horizon $t = 1,2,3, \ldots, T$, and treat the dynamic distribution network reconfiguration problem as a sequential decision-making problem: at each time t in the horizon, the algorithm select the optimal configuration with respect to the current and future time steps $t, t+1, t+2, \; \ldots$. The criterion of being optimal is to minimize:

$$\sum_{t}\sum_{l} r_l I_{lt}^2 + w|\alpha_{lt} - a_{lt-1}| \tag{2}$$

Note the time-varying variables have been indexed by time t.

In addition, operational constraints such as voltage magnitude must be enforced. The researchers consider the voltage magnitude constraint

$$V^{min} < v_{it} < V^{max} \; \forall i, \forall t \tag{3}$$

Where v_{it} is the voltage magnitude of node i at time t; V^{max} and V^{min} are the operational limits. Another operational constraint is the network radiality constraint, that is, the network configuration determined by all the open/closed status $\alpha_{1t}, \alpha_{2t}, \ldots$ at any time must be radial. This constraint is informally written as

$$\{\alpha_{1t}, \alpha_{2t}, \ldots\} \in Radial \; \forall t \tag{4}$$

The problem of dynamic distribution network reconfiguration can be summarized as:

$$
\begin{gathered}
min \sum_{t} \sum_{l} r_l I_{lt}^2 + w|\alpha_{lt} - a_{lt-1}| \\
s.t. V^{min} < v_{it} < V^{max} \; \forall i, \forall t \\
\{\alpha_{1t}, \alpha_{2t}, \dots\} \in Radial \; \forall t
\end{gathered}
\tag{5}
$$

Problem (5) can be solved by mathematical programming algorithms or heuristic methods. However, since the network parameters are unknown to most electric utilities, variables involved in the problem such as v_{it} and I_{lt}^2 cannot be calculated. Nevertheless, (5) can still be (approximately) solved by relying on the distribution network historical operational data and machine-learning algorithms. These are described in the next two sections.

Operational Data

The researchers assumed an operational historical database was available to the electric utilities, which stored operational information for a period of time (that is, half a year or a year). The database must contain the following fields in order for the machine-learning algorithm to infer useful patterns for effective reconfiguration. First, the demand and DER output shall be available for all time steps. p_{it} and q_{it} are used to denote the net real and reactive power injection at node i time t, that is, $p_{it} + jq_{it}$ is the DER output subtracted from the demand. The second field is the switch status α_{it} for a RCS and time steps. The third field is the total injected power from the substations. The total network resistive loss information is obtained by summing net real power injections at all nodes including the substation:

$$
\sum_{l} r_l I_{lt}^2 = p_t^{sub} + \sum_{i} p_{it} \tag{6}
$$

Finally, voltage magnitude measurement v_{it} can be included, if available. The notation $p_t = [p_{1t}, p_{2t}, \dots]$ $q_t = [q_{1t}, q_{2t}, \dots]$ $\alpha_t = [\alpha_{1t}, \alpha_{2t}, \dots]$ $v_t = [v_{1t}, v_{2t}, \dots]$ is used to denote the collection of all variables in the network at time t. The historical data $\{p_t, q_t, p_t^{sub}, \alpha_t, v_t\}$ will be formatted and used by the reinforcement learning (RL) algorithm. The researchers discuss the RL algorithm in the next two subsections.

Dynamic Distribution Network Reconfiguration as a Markov Decision Process

In this subsection, the researchers construct the dynamic distribution network reconfiguration problem as a Markov decision process (MDP) [8]. MDP is a standard mathematical language for defining stochastic sequential decision-making processes and describing reinforcement-learning algorithms. An MDP consists of the following elements (S, A, P, r, γ, T), which consist of a set of states S, a set of actions A, a state transition probability $P(s'|s, a) \forall\, s', a \in S, a \in A$; a reward function $r(s, a): S \times A \mapsto R\ \forall s \in S, a \in A$, a discount factor $\gamma \in [0,1]$, and a time horizon T. In an MDP, an agent selects an action $A_t \in A$ based on the environment's state $S_t \in S$ at each discrete time step t. After that the agent receives a numerical reward $R_{t+1} = r(S_t, A_t)$ and the environment's state will advance to S_{t+1} according to the state transition probability $P(S_{t+1}|S_t, A_t)$. The process terminates when $t = |T|$ and $S_{|T|}$ are a terminal state.

To describe the dynamic DNR problem in the language of MDP, the project team defined the state $s \in S$, action $a \in A$, and the reward function $r(s, a)$. The state S_t corresponds to the status of the distribution network, which includes the current topology configuration, loads, and global times. In the previously established terminologies, this means $S_t = [p_t, q_t, \alpha_{t-1}, t]$. The action A_t is represented by changing the configuration of the distribution network in S_t. That is, $A_t: \alpha_{t-1} \mapsto \alpha_t$. The reward R_{t+1} is a numerical measure of how good the reconfiguration action A_t was, in terms of minimizing network loss while keeping the number of switching actions small. It is defined as

$$r(s_t, a_t) = -p_t^{sub} - \sum_i p_{it} - w \sum_l |\alpha_{lt} - a_{lt-1}| \tag{7}$$

ol

This finishes the construction of the MDP, though several remarks follow.

1. The injection-patterns time series p_t, q_t might not be strictly Markovian. Nevertheless, this definition of S_t will still be used because the algorithms that will be discussed remain applicable even if the Markovian property is slightly violated in practice.
2. The action space so defined would include all possible radial configurations of the network. However, many of them are infeasible in terms of grid safety operations. In this project, researchers reduce the action space $\boldsymbol{A}$ to include only those configurations that appeared in the historical data set. This allows the agent to avoid selecting unacceptable network configurations, but it also limits potential discovery of optimal control policies.

During the distribution-network reconfiguration process, the nodal voltages must always stay within an allowable range. In this project, researchers incorporate a penalty for constraint violation as part of the reward function, and use $\boldsymbol{r}(\boldsymbol{S}_t, \boldsymbol{A}_t, \boldsymbol{\lambda})$ as the final reward function:

$$\begin{aligned} r(\boldsymbol{S}_t, \boldsymbol{A}_t, \lambda) = r(\boldsymbol{S}_t, \boldsymbol{A}_t) \\ - \lambda \sum_i [\max(0, v_{it} - V^{max}) + \max(0, V_{min} - v_{it})] \end{aligned} \tag{8}$$

The summation on the right-hand side of (8) is the amount of voltage violation. λ is a constant that controls the relative contribution of the constraint violation to the overall reward. The final constructed MDP replaces the reward in (7) with (8). The researchers describe the algorithm to solve the constructed MDP representing the dynamic distribution-network reconfiguration problem.

Q Learning

In this subsection, the researchers describe an off-policy reinforcement learning (RL) algorithm, known as deep Q learning [9], to solve the dynamic distribution network reconfiguration problem. Off-policy RL algorithms are a class of RL that learns and improves its control policy independently of actions taken by the agent. Therefore, they are suitable for this project since they can be adapted to learn from historical and operational data rather than from simulation. The latter is infeasible if accurate physical models are not available.

Standard tabular Q-learning algorithms update the action-value function iteratively:

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha[R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)\,] \tag{9}$$

$Q(S_{t+1}, a) = 0$ if S_{t+1} is a terminal state. The update (9) converges to the optimal action-value function Q^* under the condition that all state-action pairs (s, a) continue to be updated. Once the optimal action-value functions are learned, the optimal control policy can be found by:

$$\pi^*: S_t \mapsto \operatorname*{argmax}_a Q^*(S_t, a) \tag{10}$$

However, it is infeasible to directly apply Q-learning to the dynamic DNR problem because of the high dimensional and continuous nature of the state space. To deal with this state space, researchers parameterized and approximated the Q table (action value function) by a neural network $Q(s, a, \theta)$ [9], where θ are the neural network parameters. The action value function approximation is then with respect to the finite dimensional parameter θ instead of the original infinite dimensional Q table. Nonetheless, the neural network approximation brings an instability or even divergence problem to the learning process [9]. One cause of this divergence is the high correlations between the action values $Q(S_t, A_t)$ and the target values $R_{t+1} + \gamma \max_a Q(S_{t+1}, a)$. To stabilize the learning process, the researchers adopted a target network architecture [10] where the target values were computed by a separate neural network $Q(s, a, \theta^-)$, where θ^- are the target Q network parameters, which are only updated every C steps by $\theta^- \leftarrow \theta$. In addition to the target network architecture, the researchers also adopted the memory replay mechanism [9]. That is, the researchers stored past operational data (experiences) $e_t = (S_t, A_t, R_{t+1}, S_{t+1})$ in a memory data set $D_H = \{e_1, \cdots, e_H\}$, and randomly sampled a subset of (non-consecutive) experiences $B = \{e_t, \dots\} \subset D_H$ to update the neural network parameters.

In the dynamic network reconfiguration problem, the memory data set D_H was initially set to the network historical operational data, as described in the previous subsection; during the online application, the data set D_H will be continuously updated by the new data. As a result, the Q learning with function approximation minimizes the loss function $L(\theta)$ with respect to the parameters θ over all stored experiences:

$$L(\theta) = E\frac{1}{|B|}\sum_{e\in B}[r + \gamma \max_{a'} Q(s', a', \theta^-) - Q(s, a, \theta)] \;\; B \sim D_H \tag{11}$$

with $Q(s', a', \theta^-) = 0$ if s' is a terminal state. Once the parameters θ are learned, the network reconfiguration control policy will be given by:

$$\pi^\theta : S_t \mapsto \operatorname*{argmax}_a Q(S_t, a, \theta) \tag{12}$$

However, existing deep-reinforcement learning algorithms are typically sample-inefficient. That is, they require many samples to develop a good control policy. This means that the historical data D_H had to be large enough for an RL agent to learn a good value function (11). To improve the performance of the deep reinforcement learning algorithm for the dynamic-distribution network-reconfiguration problem, the researchers proposed an innovative technique that generates reliable, synthetic operational experience data from historical and operational data sets. This is described in the next subsection.

Operational Data Augmentation

This subsection describes the synthetic operational-experience generation (operational data augmentation) process. The researchers proposed a three-step algorithm to create a set of synthetic operational experiences $\tilde{D}_{\tilde{H}} = \{\tilde{e}_1, \cdots, \tilde{e}_{\tilde{H}}\}$ where $\tilde{e}_t = (\tilde{S}_t, \tilde{A}_t, \tilde{R}_{t+1}, \tilde{S}_{t+1})$, $\tilde{S}_t = [\tilde{p}_t, \tilde{q}_t, \tilde{\alpha}_{t-1}, t]$, $\tilde{A}_t : \tilde{\alpha}_{t-1} \mapsto \tilde{\alpha}_t$, and $\tilde{R}_{t+1} = \tilde{r}(\tilde{S}_t, \tilde{A}_t, \lambda)$. The steps are (1) synthesizing the injection time series $\{\tilde{p}_t, \tilde{q}_t\}$, (2), generating the network configuration $\tilde{\alpha}_t$ at each time step, and (3), and estimating the corresponding reward values $\tilde{r}(\tilde{S}_t, \tilde{A}_t, \lambda)$ for the data created in steps 1 and 2. Step 1 takes the historical-load time series and outputs a new one. For example, direct-historical injection data or a load-time series model using historical data can be used. In step 2, the researchers generate a sample path $\tilde{\alpha}_t$ from a stochastic process defined on the sample space A. In step 3, $\tilde{r}(\tilde{S}_t, \tilde{A}_t, \lambda)$ was estimated for each time step. The algorithms for estimating the network losses and voltage magnitudes are described here.

Two sets of regression models are trained on the historical data to estimate total network loss and nodal voltage magnitudes, respectively. For both sets of regression models, the input variables are the injection patterns and the network configurations. After the training, the reward $\tilde{r}(\tilde{S}_t, \tilde{A}_t, \lambda)$ can then be calculated based on the out-of-sample prediction of the regression models applied to the synthesized data points $\tilde{S}_t, \tilde{A}_t$. Inaccurate rewards in training data can hurt the learning process. Therefore, the researchers determined that if the estimated rewards were reliable they discarded the ones with high uncertainty. In this project, the Gaussian process (GP) [11] was used as the regression model to learn both the estimated values and their uncertainties. The theory of GP is summarized here.

Gaussian Process

In the GP setting, the target y and the input vector x are modeled by the relationship $y = f(x) + \epsilon$ where ϵ represents the observation noise and is typically a zero mean Gaussian with variance σ_ϵ^2; f is a Gaussian process $f(x) \sim GP(m(x), k(x, x'))$. If the mean function $m(x)$ covariance function $k(x, x')$ and σ_ϵ^2 are known, then the probability distribution of any data $P(y|x)$ can be evaluated and the uncertainty represented by the variance of $P(y|x)$. Typically, the mean and covariance functions are in some parametric families. For example, the constant-zero mean function and the squared-exponential-covariance function are given by:

$$m_{\theta_M}(x) = 0 \; k_{\theta_K}(x, x') = A^2 \exp\left(-\frac{||x - x'||_2^2}{2\ell^2}\right) \tag{13}$$

ty

In this example, $\theta_M = \emptyset$ and $\theta_K = \{A, \ell\}$. The researchers used the above-zero constant mean function and the squared-exponential-covariance function in this project. The parameters can be estimated by marginalizing the Gaussian process onto the training data points before performing maximum likelihood estimations of the parameters on this marginal distribution. Let the estimated parameters be $\hat{\theta}_M$, $\hat{\theta}_K$, and $\hat{\sigma}_M^2$, the posterior distribution of a testing instance x^* is again Gaussian, with the conditional mean and variance:

$$\begin{aligned} \mu(y^*|x^*) &= m_{\hat{\theta}_M}(x^*) + \Sigma_{x^*X}(\Sigma_{XX} + \hat{\sigma}_M^2 I)^{-1}(Y - m_{\hat{\theta}_M}(X)) \\ \sigma^2(y^*|x^*) &= k_{\hat{\theta}_K}(x^*, x^*) + \hat{\sigma}_M^2 + \Sigma_{x^*X}(\Sigma_{XX} + \hat{\sigma}_M^2 I)^{-1}\Sigma_{Xx^*} \end{aligned} \tag{14}$$

Now, $\mu(y^*|x^*)$ and $\sigma^2(y^*|x^*)$ represent the estimated target and its uncertainty. In the dynamic-distribution network reconfiguration problem, each x represents an injection pattern and a radial configuration and each y represents the corresponding network loss or a voltage magnitude. If the uncertainty of the target estimate $\sigma^2(y^*|x^*)$ is larger than some threshold, then the synthetic data $(x^*, \mu(y^*|x^*))$ will be discarded. In this project, the threshold is heuristically set to three times the standard deviation of all testing data points.

CHAPTER 3: Project Results

Key Findings for Three-Phase Optimal Power Flow

The proposed chordal-based convex iteration algorithm with greedy grid partition scheme is implemented in MATLAB script. Simulations are conducted on the IEEE 4-bus, 10-bus, 13-bus, 34-bus, 37-bus, 123-bus, and 906-bus three-phase test feeders. The simulation results demonstrate that the proposed algorithm achieves better performance in terms of optimality, feasibility, and scalability.

Optimality and Feasibility

To illustrate the optimality and feasibility of solutions under the proposed algorithm, a comparison of the solutions obtained from traditional methods, including the Powell, interior-point, and the proposed convex-iteration method appears in Table 1.

As shown in Table 1, the proposed convex-iteration approach achieves lower objective values on 11 out of 14 test scenarios. Traditional methods arrive at the same solution as the proposed convex iteration method on the other three test scenarios. As the size of the test feeder increases, it becomes more difficult for traditional methods to match the performance of the proposed convex-iteration algorithm.

To illustrate the global optimality and feasibility of the proposed algorithm, another comparison of solutions derived from the SDP relaxation method and the proposed convex-iteration method with the default setting is shown in Table 2.

It can be seen in Table 2 that the SDP relaxation method does not yield a rank-one solution by directly removing the rank constraint. The star symbol, $*$, represents the highest rank among all partitioned areas. The high-rank solutions do not have any physical meaning. In most cases, the solution of the SDP relaxation method provides a lower bound of the original non-convex optimization problem. On the other hand, the proposed chordal-conversion-based convex iteration algorithm always produces a rank-1 solution, which is the global optimum.

Table 1: Comparison of Traditional Methods and the Convex Iteration With Different Prices for DERs

Test System	Prices of Three Phases ($/KWh)	Objective Value ($/hour)		
		Powell	Interior Point	Convex Iteration
4-bus test feeder	1/0.5/0.2 0.9/0.45/0.18	3121.9 3091.9	3121.9 3091.9	3121.9 3086.9
10-bus test feeder	1/0.3/0.6 0.8/0.24/0.48	1229.2 1191.4	1229.2 1191.4	1229.1 1191.3
13-bus test feeder	0.6/0.3/1 0.48/0.24/0.8	2345.4 2290.2	2345.4 2290.2	2345.4 2290.2
34-bus test feeder	1/0.9/0.8 0.9/0.81/0.72	832.7 816.5	832.7 816.5	830.8 815.4
37-bus test feeder	0.6/0.3/1 0.54/0.27/0.9	1740.3 1675.9	1740.3 1675.9	1739.5 1675.4
123-bus test feeder	1/0.3/0.6 0.8/0.24/0.48	2414.6 2205.6	2414.5 2205.6	2413.6 2205.0
906-bus test feeder	0.6/0.7/0.5 0.54/0.63/0.45	38.4 37.9	38.3 37.9	38.2 37.7

kWh = kilowatt-hours

Source: Wei Wang and Nanpeng Yu, "Chordal Conversion based Convex Iteration Algorithm for Three-phase Optimal Power Flow," IEEE Transactions on Power Systems, vol. 33, no. 2, March 2018.

Table 2: Comparison of the SDP Relaxation Method and the Convex Iteration Method With Different Prices for Three Phases

Test System	Method	Rank of Solution	Objective Value ($/hour)
4-bus test feeder	SDP relaxation	3	3085.6
	Convex iteration	1	3121.9
10-bus test feeder	SDP relaxation	7	1216.3
	Convex iteration	1	1229.1
13-bus test feeder	SDP relaxation	3	2319.5
	Convex iteration	1	2345.4
34-bus test feeder	SDP relaxation	6*	831.8
	Convex iteration	1	830.8
37-bus test feeder	SDP relaxation	1	1739.5
	Convex iteration	1	1739.5

Test System	Method	Rank of Solution	Objective Value ($/hour)
123-bus test feeder	SDP relaxation	6*	2413.6
	Convex iteration	1	2413.6
906-bus test feeder	SDP relaxation	6*	38.2
	Convex iteration	1	38.2

Source: Wei Wang and Nanpeng Yu, "Chordal Conversion based Convex Iteration Algorithm for Three-phase Optimal Power Flow," IEEE Transactions on Power Systems, vol. 33, no. 2, March 2018.

Table 3: Comparison of the Penalized SDP Method and the Convex-Iteration Method With Different Prices for Three Phases

Test System	Method	Eig2/Eig1	Power injection error (kW)
4-bus test feeder	Penalized SDP	9.1×10^{-9}	5.6×10^{-3}
	Convex iteration	2.6×10^{-9}	3.9×10^{-3}
10-bus test feeder	Penalized SDP	7.7×10^{-7}	5.2×10^{-3}
	Convex iteration	2.2×10^{-9}	6.8×10^{-3}
13-bus test feeder	Penalized SDP	3.8×10^{-7}	0.2208
	Convex iteration	3.2×10^{-9}	0.0629
34-bus test feeder	Penalized SDP	1.2×10^{-5}	3.24
	Convex iteration	6.0×10^{-8}	2.41
37-bus test feeder	Penalized SDP	3.0×10^{-6}	1.54
	Convex iteration	3.0×10^{-6}	1.54
123-bus test feeder	Penalized SDP	2.8×10^{-5}	13.21
	Convex iteration	$12. \times 10^{-8}$	1.21
906-bus test feeder	Penalized SDP	5.1×10^{-5}	6.7
	Convex iteration	6.0×10^{-8}	2.3

Source: Wei Wang and Nanpeng Yu, "Chordal Conversion based Convex Iteration Algorithm for Three-phase Optimal Power Flow," IEEE Transactions on Power Systems, vol. 33, no. 2, March 2018.

At last, a comprehensive comparison between the penalized SDP method and the proposed convex-iteration algorithm was conducted. The comparison results are shown in Table 3: Comparison of the Penalized SDP Method and the Convex-Iteration Method With Different Prices for Three Phase Table 3. For the IEEE 4-bus, 10-bus, and 13-bus test feeders, the comparison is performed without graph partition. Although the penalized SDP method did obtain a rank-one solution, the ratio of the second-largest eigenvalue of matrix X to its largest eigenvalue is much larger than that of the proposed convex-iteration method. Moreover, as shown in Table 4, the power-injection error obtained from SVD of the rank-one solution of the penalized SDP method is much larger than that of the proposed convex-iteration method. For IEEE 34-bus, 123-bus, and 906-bus test feeders, the penalized SDP method fails to find a rank-one solution. The power-injection error under the penalized SDP method is also much larger than that of the proposed convex-iteration method.

Scalability

As shown in Table 4, computation times of the three-phase OPF problems on all seven IEEE test feeders are within two minutes using the entry level Dell workstation. The combination of the chordal-based conversion technique and the greedy grid partition scheme made the proposed algorithm computationally efficient.

Table 4: Scalability of the Proposed Algorithm

Test System	Computation time (s)	Number of Iteration	Number of Non-zero Elements	Rank of Solution
4-bus	0.373	4	2.95×10^4	1
10-bus	12.127	29	2.53×10^4	1
13-bus	8.714	16	3.61×10^5	1
34-bus	4.161	3	1.25×10^6	1
37-bus	3.261	1	2.06×10^6	1
123-bus	27.182	3	4.93×10^6	1
906-bus	79.799	3	1.32×10^6	1

Source: Wei Wang and Nanpeng Yu, "Chordal Conversion based Convex Iteration Algorithm for Three-phase Optimal Power Flow," IEEE Transactions on Power Systems, vol. 33, no. 2, March 2018.

Key Findings for Data-Driven Volt-VAR Control

The performance of the proposed method and the benchmarking algorithms was validated with the IEEE 4-bus and 13-bus test feeders. The results showed that the constrained-policy optimization algorithm can achieve near-optimal solutions with negligible voltage violations. Compared with the conventional optimization-based approach, the proposed reinforcement-learning algorithm is better suited for online VVC tasks where accurate and complete distribution-network models are not available.

Optimality and Constraint Satisfaction

The model predictive control (MPC) based optimization algorithm was chosen as the first benchmark. The control horizon was 24 hours. The ARIMA model was used to forecast the load during the control horizon. At each rolling step of MPC, a mixed-integer conic programming problem was solved. Two optimization packages, MOSEK and GUROBI were used to solve the mixed-integer conic programming problem. The second benchmark was set up by replacing the load forecast with actual load data in the MPC framework. The third benchmark represents the baseline where all switching devices were kept at their initial positions. The trust region policy optimization (TRPO) algorithm, which is a reinforcement-learning algorithm for MDP problems, was also implemented for comparison purposes. To be applicable for the VCC problem, the voltage violation was treated as a penalty term in the reward function.

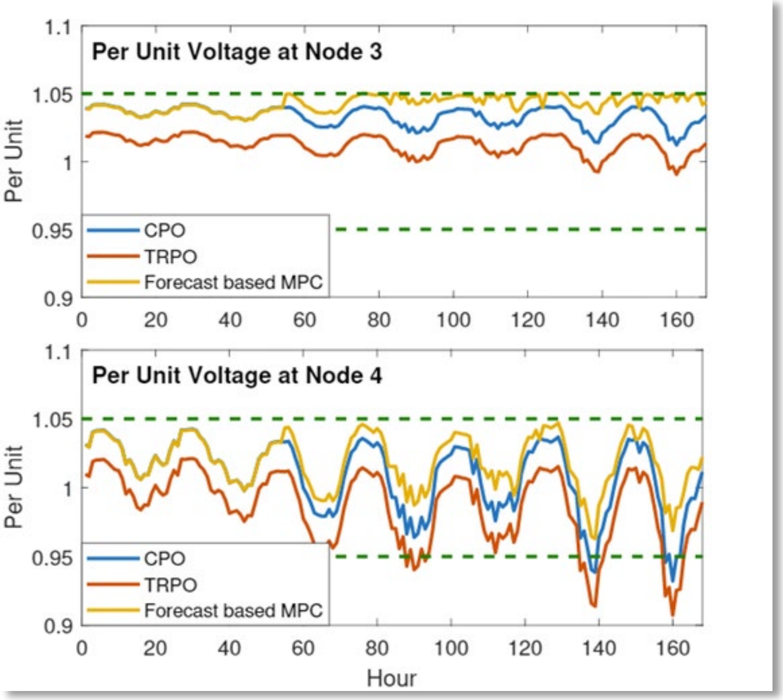
The total operation cost (OC), the number of tap changes (# of TC), the number of voltage violations (# of VV), and the accumulated per-unit voltage violation (AVV) over the test week are recorded in Table 5 for all the reinforcement algorithms and the benchmark algorithms. The operation cost includes the costs associated with line losses and tap changes. The accumulated per-unit voltage violation was calculated as the sum of voltage magnitude deviation across all the network nodes when the nodal voltage was out of operational limits.

Table 5: Performance Comparison of VVC Algorithms

	Algorithms	OC ($)	# of TC	# of VV	AMV (per unit)
Bus 4 Test Case	Baseline	150.13	0	91	2.748
	MPC(Actual)	111.44	18	0	0
	MPC (Forecast)	111.89	20	0	0
	CPO	115.01	9	5	0.044
	TRPO	120.05	3	16	0.286
Bus 13 Test Case	Baseline	77.88	0	268	2.673
	MPC(Actual)	58.05	6	0	0
	MPC (Forecast)	58.44	6	0	0
	CPO	58.92	6	0	0
	TRPO	61.29	3	2	0.0004

Source: Wei Wang, Nanpeng Yu, Jie Shi and Yuanqi Gao, "Volt-VAR Control in Power Distribution Systems with Deep Reinforcement Learning," IEEE SmartGridComm, pp. 1-7, 2019.

Figure 8: Comparison of Voltage Profiles on the 4-Bus Test Feeder



Source: Wei Wang, Nanpeng Yu, Jie Shi and Yuanqi Gao, "Volt-VAR Control in Power Distribution Systems with Deep Reinforcement Learning," IEEE SmartGridComm, pp. 1-7, 2019.

As shown in Table 5, the CPO algorithm is capable of achieving a near-optimal operational cost and is nearly constraint-satisfying. The CPO algorithm yields a lower operation cost compared with the TRPO algorithm. The per-unit voltages at node 3 and 4 of the 4-bus test feeder are shown in Figure 8. The voltage solutions at node 3 of the MPC-based approach with forecasted load hit the upper bound a few times. This is common for optimization approaches as the optimal solutions are likely to be boundary points. By following the CPO algorithm, the voltage profile sat node 4 stayed in bounds nearly all the time except for five minor violations. The CPO algorithm outperformed the TRPO algorithm by approximately satisfying voltage constraints all the time.

Table 6: Computation Times for VCC Algorithms

	Algorithms	Average Time (s)	Maximum Time (s)
Bus 4 Test Case	MPC (GUROBI)	10.43	90.28
	MPC (MOSEK)	346.80	3904.22
	TRPO/CPO	<10-3	<10-3
Bus 13 Test Case	MPC (GUROBI)	4.69	8.57
	MPC (MOSEK)	53.83	328.98
	TRPO/CPO	<10-3	<10-3

Source: Wei Wang, Nanpeng Yu, Jie Shi and Yuanqi Gao, "Volt-VAR Control in Power Distribution Systems with Deep Reinforcement Learning," IEEE SmartGridComm, pp. 1-7, 2019.

The average and the maximum computation times of the MPC-based algorithms with different solvers and policy gradient methods to determine the tap positions at each hour are provided in Table 6. Without parallel computing (MOSEK), the computation time of the MPC-based algorithm could exceed one hour in the worst case on an entry level DELL desktop. On the other hand, once trained, reinforcement-learning methods have a much faster execution speed that makes them suitable for online applications. Moreover, the MPC-based algorithms require accurate and complete topology model and parameters of the distribution network, which are not often available.

Key Findings for Reinforcement-Learning-Based Distribution Network Reconfiguration

For reinforcement-learning-based distribution network reconfigurations, key findings are summarized.

1. Operational experience augmentation can help improve reinforcement-learning algorithm performance. Specifically, the Gaussian-process model provides accurate and reliable reward value estimates.
2. The proposed reinforcement-learning algorithm for distribution-network reconfiguration reduces network resistive loss. The algorithm does not require accurate network parameter information to perform this reconfiguration, but instead learns a useful control policy from network historical data sets.
3. The proposed algorithm achieves relatively consistent results without extensive tuning of hyperparameters.

Researchers will discuss each of the findings in detail in the next subsections, though the experimental setup will be presented first.

Experiment Set-Up

Simulation Environment

To validate various concepts in this project, the 16-bus distribution test feeder [12] was used as the distribution network. The line impedances, remotely controllable switches, reference voltages, and the complex power base (MVA) were unchanged. The historical data were obtained in stages. First, the loads/DGs data were taken from the hourly smart meter data of a group of residential and commercial utility customers served by a 12 kV distribution feeder in Southern California. The length of the data was 26 weeks. Each nodal injection was set to be the aggregated consumption of a group of randomly selected customers. The researchers assumed a constant power factor for each node. Next, 83 medium-to-low line loss configurations were selected from all the radial configurations as the action space; a sample path of 26 weeks with hourly granularity was then generated from a Markov chain defined on those 83 configurations with transition probability $p_{ii} = 0.9$ and $p_{ij} = p_{ik}$. Finally, the power-flow solution was found in the simulation environment for all hours and total line losses as well as the voltage magnitude measurements at bus 7, 12, and 16 of the network were recorded to form the historical operational data set.

Set-Up of Implementation

This subsection supplements some details on the implementation of the algorithms described in the previous subsections. The neural network was feed forward with the feature encoded input $\phi(S_t) = \phi(p_t, q_t, \alpha_{t-1}, t)$. The encoding of the nodal-injection vector was the same as per-unit values; the encoding of the network configuration α_{t-1} is a sequence of on-value/off-value numbers. If the branch was closed, the number at the position corresponding to that branch was set to on-value; it was otherwise set to off-value. 0.2 for on-value and 0.0 for off-value were used in this project. The time index t in the state definition was encoded by a linearly spaced single number ranging from 0 to 1, with 0 representing the initial hour of a week and 1 representing the final hour of the week. The number of neural network output was equal to the number of elements in the action space, each corresponding to an action. The value of each of the outputs was the Q value of that action and the input state encoding so that the reward value and neural network initial weight could be in compatible ranges. The original per-unit reward value was multiplied by a factor of 50.

Researchers used the same encoding of feature and target variables for both reinforcement learning algorithms and the Gaussian process used in the data-augmentation method.

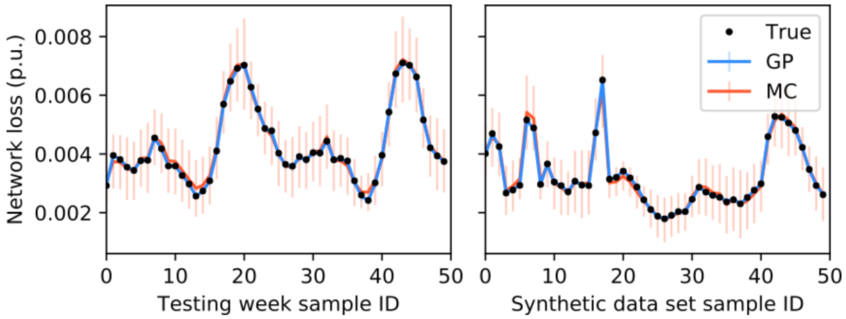
In the Gaussian process model, the covariance matrix was inverted to perform the inference. However, the matrix was not guaranteed to be numerically invertible. In this project, a small constant (1×10^{-6}) was added to the diagonal of the covariance before inversion. This helped to improve the numerical stability without overly affecting the results.

Operational Experience Augmentation

This subsection validates the quality of the synthetic operational experience data generated by the proposed Gaussian-process-based model. Researchers first demonstrated that the Gaussian process is more accurate than another popular nonlinear regression model: the Monte Carlo dropout neural network [13].

Of the 26 weeks of historical data, the first 25 weeks were chosen as training data and the last week's data were chosen as testing data. Researchers then created a 25-week synthetic operational data set $\widetilde{D}_{\widetilde{H}}$ as follows. First, the researchers generated a 25-week sample path of network configurations from a Markov chain, defined on configurations that appeared in the training data set, with transition probability $p_{ii} = 0.8$ and $p_{ij} = p_{ik}$. Researchers then estimated network losses for this new sequence of configurations under the injection patterns of the first 25 weeks of historical data. For network loss estimates, the researchers compared the Gaussian-process model with the Monte Carlo dropout neural network. Both the Gaussian-process model and the Monte Carlo dropout model were trained with the first 25 weeks of historical operational data. The researchers applied the trained model to the 1-week testing data set and the 25-week synthetic operation experience data set shows the performance of network loss predictions for the two models under 50 samples of the testing and synthetic data sets. As shown in Figure 9, compared with the MC dropout model the GP model more accurately predicted network losses.

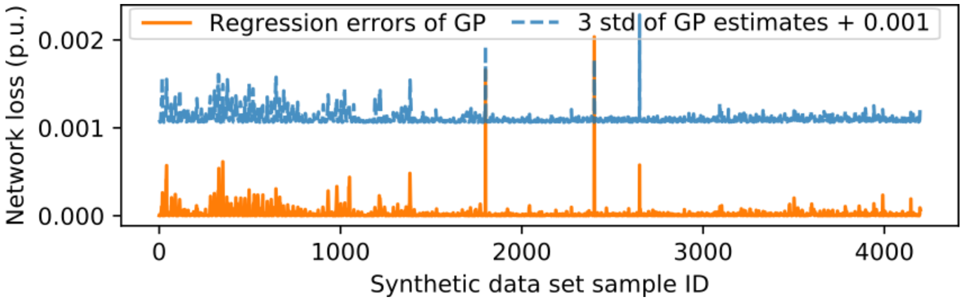
Figure 9: Performance of Out-of-Sample Predictions for Network Losses



Source: Wei Wang, Nanpeng Yu, Jie Shi and Yuanqi Gao, "Volt-VAR Control in Power Distribution Systems with Deep Reinforcement Learning," IEEE SmartGridComm, pp. 1-7, 2019.

Although GP models produce fairly accurate predictions, they occasionally lead to large errors for some network configurations and injection patterns, as shown by the orange curve in Figure 10. Fortunately, the uncertainty estimates of the GP model represented by the blue curve correlate very well with the estimation error. This suggests that the proposed strategy of removing the samples with large uncertainty estimates improves the quality of augmented operational data.

Figure 10: Regression Errors Versus Uncertainty Estimates of the GP Model

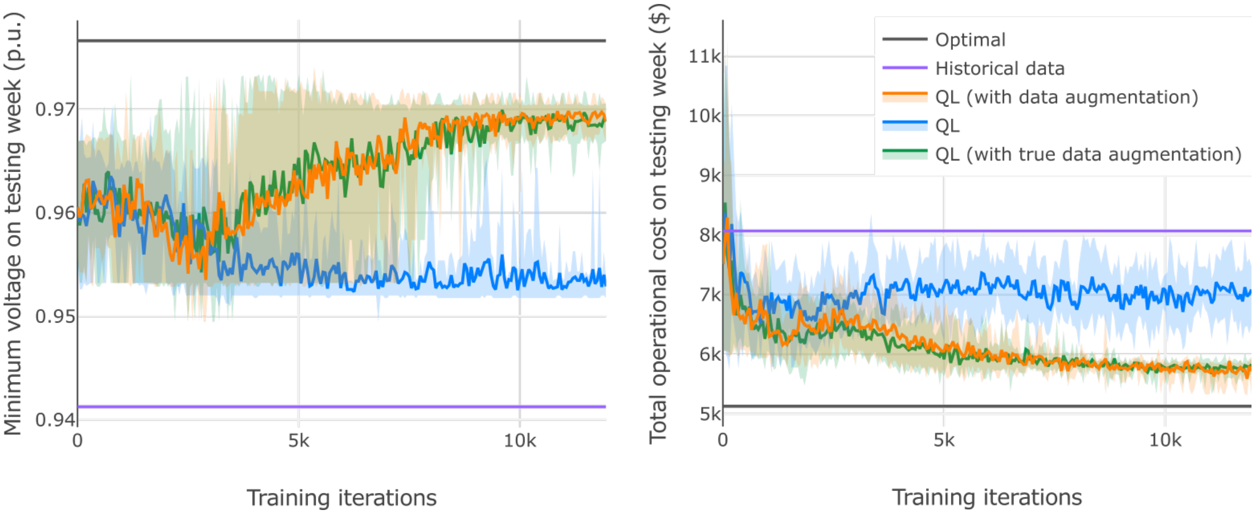


Source: Wei Wang, Nanpeng Yu, Jie Shi and Yuanqi Gao, "Volt-VAR Control in Power Distribution Systems with Deep Reinforcement Learning," IEEE SmartGridComm, pp. 1-7, 2019.

Performance of Deep Q-Learning Algorithms

In this subsection, researchers compared the performance of three deep Q-learning algorithms with two benchmarks. In the first benchmark, a global optimal solution of the dynamic distribution network reconfiguration problem was obtained through dynamic programming with perfect knowledge of the network parameters and future injection patterns. The second benchmark simply used historical network configurations in the data set. The first Q-learning algorithm was trained using only historical data. The second Q-learning algorithm was trained with both historical and synthetic experiences, and network losses were estimated based on the GP model. The third deep Q-learning algorithm was trained with both historical and synthetic operational experiences where the network losses were obtained from power-flow studies that assumed perfect knowledge of network parameters. The researchers divided the 26-week historical data set into a 25-week training data set and a 1-week testing data set. The 25-week synthetic operational experience data set was generated in the same way described in the previous subsection. During the training iterations, researchers periodically saved the parameters of the neural network and tested its performance on the testing data set. Some hyperparameters were given as follows: the neural network was a feed-forward, 2-layer net with 600 hidden neurons and 83 outputs; the activation function was ReLu; the optimization algorithm was chosen as Adam; the mini-batch size used in the stochastic gradient descent was 64; the discount factor was chosen as 0.95; and the update step C was set to 30. The performance of the three Q-learning algorithms and two benchmarks is shown in Figure 11.

Figure 11: Performance of Q-Learning



Source: Wei Wang, Nanpeng Yu, Jie Shi and Yuanqi Gao, "Volt-VAR Control in Power Distribution Systems with Deep Reinforcement Learning," IEEE SmartGridComm, pp. 1-7, 2019.

The left subfigure shows the minimum-voltage magnitude over all metered nodes and all hours. The right subfigure shows the total operational cost. For the three deep Q-learning algorithms, the average, the 10th, and the 90th percentiles of the results from 10 independent runs are shown. Compared with network configurations in the testing week of the historical data, the deep Q-learning algorithm quickly learned how to reduce operational cost in a dynamic distribution-network reconfiguration problem. When the historical operational experiences were augmented with synthetic operational data, the operational cost of the deep Q-learning algorithm was further reduced, and the minimum-voltage magnitudes got even closer to nominal voltage values. As the learning process proceeded, the performance of the deep Q-learning algorithms with augmented operational experiences approached that of the global optimal solution. Note that the Q-learning agents achieved these results without knowing actual network parameters or future power-injection patterns. The orange curve almost coincides with the green curve, suggesting that the network losses estimated by the proposed Gaussian process model are almost as good as the power-flow solutions with perfect network parameter information.

The consistency of the Q-learning performance is discussed next. The researchers demonstrated that similar results can be obtained without extensively tuning hyperparameters, which is crucial for practical applications. Researchers demonstrated this by showing that the performance of the Q-learning algorithm is relatively consistent under different hyperparameter settings. The following combinations of hyperparameters were tested: batch size $B \in \{32,64,128,256\}$, number of hidden layers $L \in \{1,2\}$, number of hidden neurons $H \in \{300,400,500,600\}$, and number of steps the target Q network's parameters were updated $C \in \{30,60,90,120\}$. The researchers generated a

Taguchi’s orthogonal array for these hyperparameter combinations and reported the results in Table 6.

Each calculated cost represents the average of five independent runs for Q-learning with operational-data augmentation. Compared with historical operational cost and optimal cost, the operational cost of the Q-learning algorithm under different hyperparameter settings is quite consistent.

Table 7: Operational Costs With Various Hyperparameters

Original cost: $ 8066.7					**Optimal cost: $ 5128.8**				
B	**H**	**C**	**L**	**QL cost ($)**	**B**	**H**	**C**	**L**	**QL cost ($)**
32	300	30	1	5752.9	128	300	90	1	5490.6
32	400	60	1	5686.2	128	400	120	1	5647.4
32	500	90	2	5608.8	128	500	30	2	5441.0
32	600	120	2	5464.3	128	600	60	2	5396.7
64	300	30	2	5523.9	256	300	120	2	5480.4
64	400	60	2	5456.8	256	400	90	2	5487.8
64	500	90	1	5579.3	256	500	60	1	5652.2
64	500	90	1	5579.3	256	500	60	1	5652.2

Source: Yuanqi Gao, Jie Shi, Wei Wang and Nanpeng Yu, "Dynamic Distribution Network Reconfiguration Using Reinforcement Learning," IEEE SmartGridComm, pp. 1-7, 2019.

CHAPTER 4: Technology/Knowledge/Market Transfer Activities

Knowledge Transfer Activities

Knowledge gained in this project has been shared with the industry and academia through three channels. First, the project team disseminated research results via conferences and journal papers. Second, project team members delivered presentations through conferences for both researchers in academia and practitioners in the utility industry. Third, project team members visited Southern California Edison, Lawrence Livermore National Laboratory, and Pacific Northwest National Laboratory to describe and share knowledge learned in this project.

In total, the research project has so far resulted in 10 international conferences and journal papers. These publications appeared in power-industry top venues such as IEEE Transactions on Smart Grid, IEEE Transactions on Power Systems, Applied Energy, and IEEE SmartGridComm.

The Principle Investigator also delivered over a dozen presentations at conferences such as IEEE SmartGridComm, IEEE Power and Energy Society General Meeting, and IEEE PES T&D Conference and Exposition.

In particular, the PI met with the principal manager, senior managers, and engineers of Southern California Edison's advanced technology laboratory and presented the iDERMs. The SCE team strongly recommended that the research team reach out to leading software vendors such as General Electric and Siemens AG to incorporate key software modules of iDERMS into the vendors' advanced distribution management systems.

The PI also visited Lawrence Livermore National Laboratory and Pacific Northwest National Laboratory to deliver seminars. Research staff at the two national laboratories showed great interest in the iDERMS software modules.

Technology Transfer Activities

Technologies developed in this project will be shared with the utility industry through two venues. First, the open-source software modules developed in this project have been shared with the public through the project's official website. Second, the project team has been actively communicating with energy-industry software vendors such as General Electric and Siemens AG about integrating the software developed in this project into their commercial products for advanced distribution-management systems. One pathway to commercializing iDERMS is for the research team to collaborate with software vendors to advance large-scale demonstration and implementation opportunities with electric utilities.

CHAPTER 5: Conclusions/Recommendations

Conclusions

The project team successfully developed an iDERMS and achieved all three project goals. First, researchers developed a three-phase optimal power flow algorithm that coordinated operations of a large number of distributed energy resources in electricity distribution systems. Second, the researchers developed a decentralized Volt-VAR control algorithm that reduced network losses and maintained customer voltages. Third, the researchers developed a data-driven distribution network reconfiguration algorithm that reduced both network losses and outage durations and frequency.

The technologies developed in this project have drawn great interest from various stakeholders such as microgrid operators, electric utilities, and software vendors. Broad adoption of the proposed algorithms and software modules could lead to significant energy savings, reductions in greenhouse gas emissions, and greater electric system reliability.

Recommendations

To further improve applicability of the proposed technology in communication-constrained systems, further research and field tests are required to develop data-driven distribution-system control algorithms. More specifically, the proposed single-agent reinforcement learning algorithms should be extended to multi-agent reinforcement learning algorithms. This broader approach will reduce the need for frequent communication between customer distributed-energy resources and distribution-system operators. A field demonstration of iDERMS on existing distribution systems would be very valuable. This demonstration would require close collaboration with an electric utility that operates with remote-controlled devices and some sensors from SCADA and other advanced metering infrastructures.

Lessons Learned

Three critical lessons were learned through this research project. First, pushing new technology into the electric utility industry will require targeted training and education for utility workforces and buy-in from regulatory agencies. In particular, the electric utility workforce is not very familiar with machine-learning techniques. The acceptance of data-driven control modules requires that distribution-system operators have a reasonable understanding of the basics of machine-learning algorithms. Second, it is extremely important to work side-by-side with electricity utility companies when developing advanced control algorithms and software in distribution networks. Holding face-to-face meetings with utilities early in the process helped steer this project in the right direction. Feedback from the electric utilities clarified the real-world utility challenges of incorporating large amounts of distributed-energy resources into established distribution and transmission systems.

CHAPTER 6: Benefits to Ratepayers

Overview

Three decentralized control algorithms in the distribution network worked together to enhance grid reliability, lower consumer electricity costs, and improve safety. Specifically, the decentralized three-phase optimal power flow optimized distribution-system energy dispatch to reduce electricity costs. It also ensured that the system operated in security regions. The decentralized Volt-VAR control algorithm reduced peak feeder loads and prevented voltage excursions. The distribution network reconfiguration and restoration technology also enhanced distribution network reliability by anticipating unfavorable renewable and load dynamics that could cause system disturbances.

Quantitative Estimates of Potential Benefits

The proposed three-phase optimal-power-flow module on the iDERMS platform could reduce energy dispatch costs in power-distribution systems by up to 10 percent. The proposed data-driven Volt-VAR control algorithm could also reduce the distribution network losses and operational costs of voltage-regulating devices by 10 percent. A more detailed analysis of energy and cost savings appears in the project's publication [14] [15] [16] [17]. The total electric load consumed by California residents in 2018 was 285,488 gigawatt-hours (GWh). The average price paid by electric utility customers of SCE, PG&E, and SDG&E were $0.149/kWh, $0.183/kWh, $0.205/kWh in 2016. If all electric utilities adopted the proposed technologies, the potential economic savings of the proposed technology can be calculated as 10% × 285,488 GWh × $0.150/kWh = $4.282 billion. Up to 285,488 GWh × 10% = 28,549 GWh of energy could be saved. The equivalent reduction in greenhouse gas emissions is up to 13,103,991 metric tons. The technology can also be adopted by microgrid operators. A similar percentage of savings can be achieved with this proposed technology. This research has set the groundwork for greater hosting capacity analyses on electric-distribution circuits in California.

GLOSSARY AND LIST OF ACRONYMS

Term	Definition
DMS	Distribution Management System
iDERMS	Integrated Distributed Energy Resources Management System
Feeder	A component of the power distribution network, which enables power to flow from the distribution substation to each customer.
Node	Any point on a power system where the terminals of two or more circuit elements meet.
Voltage Excursion	Voltage magnitude exceeding normal operating range
Radial Network	A radial network is arranged like a tree.
Meshed Network	A network with loop(s).
Multi-agent System	A system with multiple agents.
Network Reconfiguration	Change the topology of the power distribution network.
Convex Iteration	An optimization framework with iterative convex optimization modules
DERs	Distributed Energy Resources
ADMS	Advanced Distribution Management System
VVC	Volt-VAR Control
OPF	Optimal Power Flow
DSO	Distribution System Operator
DRL	Deep Reinforcement Learning
CSAC	Constrained Soft Actor-Critic
DNR	Distribution Network Reconfiguration
RCSs	Remotely Controllable Switches
MIP	Mixed-Integer Programming
RL	Reinforcement Learning
SDP	Semi-definite Programing
SQP	Sequential Quadratic Programming
CMDP	Constrained Markov Decision Process
MDP	Markov Decision Process
CPO	Constrained Policy Optimization

REFERENCES

[1] R. 16-02-007, "Decision Adopting Preferred System Portfolio and Plan for 2017-2018 Integrated Resource Plan Cycle," http://docs.cpuc.ca.gov/PublishedDocs/Published/G000/M284/K786/284786020.PDF, 2019.

[2] S. Frank and S. Rebennack, "An introduction to optimal power flow: Theory, formulation, and examples," *IIE Transactions,* vol. 48, pp. 1172--1197, 2016.

[3] B. F. Wollenberg, Power System Operation and Control, 2001.

[4] W. Wang and N. Yu, Chordal Conversion Based Convex Iteration Algorithm for Three-Phase Optimal Power Flow Problems, vol. 33, IEEE Transactions on Power Systems, 2018, pp. 1603-1613.

[5] T. A. Short, Electric Power Distribution Handbook, Taylor & Francis, 2014.

[6] T. Haarnoja, A. Zhou, P. Abbeel and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," *arXiv preprint arXiv:1801.01290,* 2018.

[7] D. Shirmohammadi, "Service restoration in distribution networks via network reconfiguration," *IEEE Transactions on Power Delivery,* vol. 7, pp. 952-958, 1992.

[8] R. A. Jabr, R. Singh and B. C. Pal, "Minimum Loss Network Reconfiguration Using Mixed-Integer Convex Programming," *IEEE Transactions on Power Systems,* vol. 27, pp. 1106-1115, May 2012.

[9] R. S. Sutton and A. G. Barto, Reinforcement learning: An introduction, MIT press, 2018.

[10] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski and others, "Human-level control through deep reinforcement learning," *Nature,* vol. 518, p. 529, 2015.

[11] A. Nair, P. Srinivasan, S. Blackwell, C. Alcicek, R. Fearon, A. De Maria, V. Panneershelvam, M. Suleyman, C. Beattie, S. Petersen and others, "Massively parallel methods for deep reinforcement learning," *arXiv preprint arXiv:1507.04296,* 2015.

[12] C. E. Rasmussen, Gaussian processes in machine learning, Springer, 2003.

[13] C.-T. Su and C.-S. Lee, "Network reconfiguration of distribution systems using improved mixed-integer hybrid differential evolution," *IEEE Transactions on Power Delivery,* vol. 18, pp. 1022-1027, July 2003.

[14] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *International Conference on Machine Learning*, 2016.

[15] Y. Liu, N. Yu, W. Wang, X. Guan, Z. Xu, B. Dong and T. Liu, "Coordinating the Operations of Smart Buildings in Smart Grids," *Applied Energy,* vol. 228, pp. 2510-2525, 2018.

[16] W. Wang, N. Yu, Y. Gao and J. Shi, "Safe Off-Policy Deep Reinforcement Learning Algorithm for Volt-VAR Control Problems in Power Distribution Systems," *IEEE Transactions on Smart Grid,* 2020.

[17] W. Wang, N. Yu, J. Shi and Y. Gao, "Volt-VAR Control in Power Distribution Systems with Deep Reinforcement Learning," in *IEEE SmartGridComm*, 2019.

[18] Y. Gao, J. Shi, W. Wang and N. Yu, "Dynamic Distribution Network Reconfiguration Using Reinforcement Learning," in *IEEE SmartGridComm*, 2019.