DOCKETED

Docket Number:	17-EPIC-01
Project Title:	Development of the California Energy Commission Electric Program Investment Charge 2018 â€“ 2020 Triennial Investment Plan
TN #:	215870
Document Title:	Nanpeng Yu Comments: Big Data Applications in Smart Power Distribution Systems
Description:	N/A
Filer:	System
Organization:	Nanpeng Yu
Submitter Role:	Applicant
Submission Date:	2/10/2017 3:51:16 PM
Docketed Date:	2/10/2017

Comment Received From: Nanpeng Yu
Submitted On: 2/10/2017
Docket Number: 17-EPIC-01

Big Data Applications in Smart Power Distribution Systems

Additional submitted attachment is included below.

CALIFORNIA ENERGY COMMISSION
1516 NINTH STREET
SACRAMENTO, CA 95814-5512
www.energy.ca.gov

The Energy Commission is currently soliciting ideas and stakeholder input for the 2018 – 2020 EPIC Triennial Investment Plan. For those that would like to submit an idea for consideration in the 2018-2020 EPIC Triennial Plan, we ask that you complete the form below. Submittals are due by **5:00 p.m. on February 10, 2017.**

Part 1. Initiative Description and Purpose:

1. Please provide a brief description of the proposed initiative:

Initiative Title: Big Data Applications in Smart Power Distribution Systems

Initiative Description: The proposed initiative enhances the modeling, monitoring, and maintenance of electric power distribution systems with big data analytics. It enables the California power grid rapidly transform from a centralized an unidirectional power distribution network to a decentralized, bidirectional network supporting large penetrations of distributed energy resources and renewables.

Initiative Goal: The goal of the proposed initiative is to perform the critically needed research to enhance the modeling, monitoring, and maintenance of electric power distribution systems with advanced data mining techniques. Specifically, the project objectives are to 1) enhance the distribution system modeling by developing adaptive topological models and state estimators, 2) develop and validate spatio-temporal models of loads and DERs through net load disaggregation, and 3) develop a preventive maintenance toolbox to predict power distribution system component failure and degradation. Moreover, the proposed data mining techniques and software will be validated using large-scale distribution system sensor data.

Executive Summary: The U.S. power grid is rapidly transforming from a centralized and unidirectional power distribution network to a decentralized, bidirectional network supporting large penetrations of distributed energy resources (DERs). To seamlessly integrate large-scale heterogeneous DERs into the distribution grid, significant enhancements need to be made to improve the modeling, monitoring, and maintenance of electric power distribution systems. Fundamental applications such as state estimation, phase identification, net-load decomposition, and predictive maintenance are still in the early stage of development in most of the electric utility companies.

Background: Penetration of advanced sensor systems such as advanced metering infrastructure (AMI) have been increasing significantly in power distribution systems over the past few years. According to U.S. energy information administration (EIA), the aggregated AMI installation experienced a 17 times increase from 2007 to 2012. The AMI usually collects electricity usage data every 15 minute, instead of once a month. This is a 3,000 fold increase in the amount of data utilities would have processed in the past. It is estimated that the electricity usage data collected through AMI in the U.S. amount to well above 100 terabytes in 2012. To unleash full value of the complex data sets, innovative big data algorithms need to be developed to transform the way we operate and plan for the distribution system.

2. What technical and/or market barriers would the proposed initiative help overcome? For scientific analysis and tools, what knowledge gaps would the proposed initiative help fill?

Technical barriers: Most of the control and coordination algorithms for DERs depend on an accurate distribution network and phase connectivity model. Although the network connectivity model of a given network is usually quite accurate, phasing errors are extremely common. In fact, the majority of distribution utilities do not have any information about the phase connectivity of their distribution circuits. Without accurate phase connectivity model, most of the operational application and planning applications will provide incorrect results. Therefore, an accurate phase identification model is in critical need. Similarly, in order to control and manage a large numbers of DERs, the system operator needs a clear understanding of the states of the distribution systems. However, most electric utilities do not have a robust distribution system state estimation engine and instead reply on a heuristic power flow approach.

Knowledge gaps that the proposed initiative help fill

Distribution network phase connectivity identification technology: Very few methods of phase identification have been studied. The existing methods can be separated into two general approaches. The first approach uses only smart meter data and supervisory control and data acquisition (SCADA) information. This approach is not even 60% accurate at identifying phase connectivity. Moreover, none of the methods in this approach can be scaled up to provide real-time phase identification for the thousands of distribution circuits in a regional power network. In the second approach, special equipment is installed on the grid. This equipment can include micro-synchrophasors, signal generators, and discriminators. A new generation of phase identification technology should increase the estimation accuracy to 90%.

Distribution system state estimation technology: Distribution system state estimation is a relatively new topic. Very few researchers have attempted to address the state estimation problem in the electric power distribution systems. Baran et al. pointed out that the voltage and electric load measurements at customer sites can be used for state estimation in the distribution systems. A group of researchers developed a DSSE prototype utilizing smart meter and SCADA data which are assumed to be available by phase. The access to accurate AMI data can improve the state estimation results in power distribution systems. Some researchers provided a detailed discussion of three-phase distribution system modeling, measurement functions, and constraints. In terms of the DSSE algorithm design, the early work in distribution state estimation adopted the weighted least squares (WLS) method. A current-based fast decoupled state estimation algorithm was developed for distribution systems. A distribution system state estimation algorithm considering nonsynchronized smart meter data was developed by modeling the load variations. A new generation of three-phase state estimation method can significantly improve the numerical stability and computational efficiency of the existing methods. Most electric distribution utilities do not run a true state estimation in real-time operations on their distribution circuits. The main reasons are 1) the phase connectivity information and transformer to customer association information are largely missing or incorrect, 2) the low level of measurement redundancy can cause the system to be unobservable, and 3) the asynchronous measurements from smart meters make it difficult to interpret the state estimation results. In reality, most of the electric distribution utilities simply allocate the loads readings from SCADA system randomly to the three phases and perform a three-phase power flow analysis instead of a robust state estimation.

Part 2. Benefits and Impacts

3. If this initiative is successful, either fully or partially, what would be the expected impact? Who are the primary users and/or beneficiaries?

The electric power utilities would the primary users of the big data applications for enhancing the modeling, monitoring, and maintenance of electric power distribution system.

The phase identification technology from this initiative will be at least 90% accurate at identifying the phase connectivity of customers/structures in electric power distribution networks. The phase identification technology only needs smart meter data and SCADA information as inputs. The usage of network topology information and historical outage information will be purely. In addition, with the implementation of the proposed technology in PySpark, the toolbox will simultaneously identify the phase connectivities of thousands of circuits in a utility company's service territory.A cost reduction of over 50% is expected over the current commercial products because no additional sensors need to be installed in the distribution network. The accuracy improvement comes from the utilization of known information about line configuration in the network connectivity model to avoid mislabeling of customers on the same secondary feeder.

Unlike the existing approach, the proposed state estimation algorithm can model all types of distribution networks with a mix of single-phase, two-phase and three-phase laterals. The proposed DSSE also models the electrical connections among customer/smart meters, the center tapped transformers, and the secondary feeders. This is missing in previous methodologies. Regarding the solution technique for the state estimation problem, the previous algorithms rely on a tunable parameter to solve a least square problem. However, it is difficult to find an appropriate parameter for all distribution feeders. In this project, the proposed state estimation algorithm has better numerical stability and no longer needs any tunable parameter. In addition, the conventional state estimation algorithm is not appropriate for distribution system application due to the low sampling frequency and asynchronous measurements from smart meters. The proposed DSSE enables forecast-aided state estimation which incorporates the spatio-temporal load and DERs forecasts into the state estimation process.

The current models for estimating probability of asset failure in distribution network typically only consider the age. When age is the only determining factor, the probability that a replaced transformer would have failed in the next 3 years is only 18%. In the proposed preventive equipment maintenance model, additional explanatory variables will be considered which include manufacturer, voltage level, weather, and electric loading level. The proposed preventive maintenance tool will increase the probability that a replaced transformer fails in the next 3 years to over 50%. Advanced machine learning and deep learning models including random forests and recurrent neural networks will be developed to improve the forecasts for equipment failure probability. In addition, the preventive equipment maintenance toolbox introduces the concept of risk score which includes both the probability of failure and the asset failure consequence score. In other words, even if the probability of failure is not high for a distribution equipment, the preventive replacement will still take place if the consequence of the equipment failure is very high.

4. Describe what quantitative or qualitative metrics or indicators would be used to evaluate the impacts of the proposed initiative:

Network topology identification technology. The quantitative metric to evaluate the technology is the accuracy of the identification results. The successful implementation of the phase identification technology enables the electric utilities to better perform load balancing, power flow analysis, state estimation, network reconfiguration, and system restoration. The proposed technology will enable higher renewable penetration in the distribution network and enhance distribution system reliability.

State estimation technology. The quantitative metric to evaluate the technology is the accuracy of the state estimation results. An ideal system should provide robust estimation results which has less than 5% of error for reasonable measurement noise variance levels. The technology will greatly enhance the grid visibility and reliability.

Preventive Equipment Maintenance. If the equipment (e.g., transformer) is replaced only based on the age factor, then the proportion of replaced equipment that would have failed in the next 3 years is typically less than 30%. The preventive equipment maintenance toolbox will ensure that the proportion of proactively replaced equipment that would have failed in the next 3 years is above 50%. The savings from the implementation of the technology will be reduction in infrastructure cost.

5. Please provide a list of peer-reviewed references that support the responses for questions 3 and 4. Proposed initiatives that include peer-reviewed references will be given stronger consideration.

Nanpeng Yu, Sunil Shah, Raymond Johnson, Robert Sherick, Mingguo Hong and Kenneth Loparo, "Big Data Analytics in Power Distribution Systems", IEEE PES Conference on Intelligent Smart Grid Technology, Washington DC, Feb. 2015.

Wenyu Wang, Nanpeng Yu, Brandon Foggo, and Joshua Davis, "Phase Identification in Electric Power Distribution Systems by Clustering of Smart Meter Data," in 15th IEEE International Conference on Machine Learning and Applications, pp. 1-7, Anaheim, California, 2016.

Yuanqi Gao and Nanpeng Yu, "State Estimation for Unbalanced Electric Power Distribution Systems Using AMI Data," in the Eighth Conference on Innovative Smart Grid Technologies (ISGT 2017), pp. 1-5, Arlington, VA

Brandon Foggo, Nanpeng Yu, and Wenyu Wang, "Distribution Network Phase Identification through Topological Data Analysis," to appear in IEEE Transactions on Smart Grid, 2017.

Wenyu Wang, Nanpeng Yu, and Raymond Johnson "A Model for Commercial Adoption of Photovoltaic Systems," to appear in Journal of Renewable and Sustainable Energy, 2017.

6. (For technologies only) What competitive advantages does the proposed technology solution have over current benchmark technologies? If the technology is beyond the prototype stage, what strategies do you suggest to bring to scale?

Competitive advantages over current benchmark technologies.
The proposed phase identification toolbox comprises of both unsupervised machine learning algorithm and supervised machine learning algorithm. The unsupervised machine learning algorithm pushes the innovation frontier in the following ways, 1) the proposed technique utilizes the known information about line configuration in the network connectivity model to avoid mislabeling customers on the same secondary feeder, 2) the proposed method can determine the phase connections of metered customers even when the circuit has some unmetered loads, and 3) the proposed technology can identify phase connection with high accuracy in distribution circuits where the majority of the loads are connected to two-phase laterals.

The proposed data driven method of net-load disaggregation and the spatio-temporal forecast model for electric load and DERs are innovative in the following ways, 1) using the customers who do not have DERs as explanatory variables can naturally control other factors that might affect electric load, such as temperature, humidity, seasons, and hours and thus avoid the possible misspecification of those factors, 2) a feature screening method to first reduce ultrahigh dimensional explanatory variables to a moderate size can vastly speed the computation in our big-data setting and improve the accuracy of the model estimate due to the much reduced noise, 3) a penalized regression is used to simultaneously estimate the regression parameters and select important explanatory variables, 4) using a spatio-temporal model can incorporate the correlation and borrow the information across both space and time, 5) the semiparametric mean regression model can blend the effects of both linear and nonlinear effects of explanatory variables and thus reduce the model bias, and 6) the robust MM-estimate can provide more stable/robust regression estimate when the data contain outliers.

The proposed preventive equipment maintenance toolbox will lead to the following technological advancements, 1) the equipment failure forecast will be 3 times accurate than Weibull curves due to the adoption of deep learning algorithms, 2) the proactive equipment replacement will be based on both the probability of failure as well as the consequence of the equipment failure, and 3) in addition to asset age, a comprehensive list of explanatory variables will be integrated into the failure prediction module which includes manufacturer, voltage level, weather, and electric loading level.

Strategy to bring the technology to scale. A project team needs to work closely together with the distribution utility partners. As soon as the proposed technology passes the verification and validation tests in the field, the project team need to engage major utility companies to set up additional large-scale pilot demonstrations of the proposed technology. To move the proposed technology from the large-scale pilot demonstration into commercialization and deployment, the project team should form strategic alliances with both traditional power system vendors and big data vendors . For example, one commercialization path is to integrate the proposed technologies into the existing vendor products.

Part 3. Connection to Energy Commission's EPIC Framework
Energy Commission staff have developed a draft strategic framework to guide the CEC's planning and implementation of EPIC across triennial investment cycles. One of the objectives of the draft strategic framework is to communicate a consistent set of priorities for organizing current and future EPIC investments.

7. Please indicate which of the following strategic framework themes you feel the proposed initiative best fits within:
    - Advance Technology Solutions for Deep Energy Savings in Building and Facilities
    - Accelerate Widespread Customer Adoption of Distributed Energy Resources
    - Increase System Flexibility from Low-Carbon Resources
    - Increase the Cost-Competiveness of Renewable Generation
    - Create a Statewide Ecosystem for Incubating New Energy Innovations
    - Maximize Synergies in the Water-Energy-Food Nexus
    - Develop Tools and Analysis to Inform Energy Policy and Planning Decisions
    - Catalyze Clean Energy Investments in California's Underrepresented and Disadvantaged Communities

If Other, Please Specify

Enhance reliability of the electric power distribution network