Evaluating the Suitability of a Grid Topology Discovery Algorithm for Utility Deployment

Shiva Poudel†, Andrew Fisher†, Aaqib Peerzada†, Arun Veeramany†, Andrew P. Reiman†,
Vitalii Dubonos‡, Peter Jang‡, Young Ngo‡, and Amy Grice§

†Pacific Northwest National Laboratory, Richland, WA 99354, USA
‡Survalent Technology Inc., Brampton, ON L6Y 5X5, Canada
§Peninsula Light Co., Gig Harbor, WA 98332, USA

***Abstract*—This paper presents the results of a field-informed demonstration aimed at evaluating the practical suitability of a topology discovery algorithm for utility environments. We demonstrated an algorithm that uses a graph-theory-informed state estimation approach for model selection. In collaboration with Survalent and Peninsula Light Co., the algorithm was applied to real feeder models and field measurements from supervisory control and data acquisition (SCADA) and advanced metering infrastructure (AMI) systems to identify the operational topology of a power distribution system. The demonstration assessed the algorithm's performance under realistic data conditions, including sparse and noisy measurements, and examined its ability to identify the most likely network configurations. The results confirmed that the approach can effectively narrow down feasible topologies, providing operators with improved situational awareness of network status. Key lessons learned emphasize the need for systematic data validation and strategic sensor placement to enhance observability. These insights inform future deployment strategies and guide refinements for broader adoption in utility operations.**

***Index Terms*—Distribution system topology, topology discovery, state estimation, and sensor placement.**

I. Introduction

Advanced distribution management applications such as model-based state estimation, volt/VAR control, distributed energy resource management, and outage management rely heavily on maintaining an accurate grid model in the distribution system operator's control center [1]; however, due to the highly dynamic nature of distribution systems, which can undergo multiple reconfigurations in a single day [2], keeping the system topology current remains a significant challenge. Without an accurate and up-to-date representation of the network, operators risk making decisions based on outdated information, which can potentially compromise system efficiency, reliability, and resilience [3].

In transmission systems, the topology identification task is relatively straightforward due to the widespread deployment of telemetered devices, supervisory control and data acquisition (SCADA) systems, and phasor measurement units, which provide rich and high-resolution data for processing topology changes [4]. Unlike transmission systems, which benefit from extensive redundancy in their metering infrastructure (with redundancy ratios of 1.7–2.2), distribution networks are sparsely equipped with real-time measurements. Different categories for solving the topology identification problem in distribution grid are proposed that are based on 1) generalized state estimation (GSE), 2) optimization-based methods, and 3) data-driven approaches [5]. Such methods have shown promising results on IEEE test feeders; however, they often rely on unrealistic assumptions and fail to capture the practical constraints and operational challenges of real distribution grids. These limitations stem from several factors, including reliance on high-frequency voltage and current measurements from micro-PMUs, assumptions that only one switching action occurs at a time, the computational complexity of solving NP-hard mixed-integer formulations, and the need to enumerate all possible candidate topologies. Furthermore, changes in grid structure and function, as well as changes in grid requirements, make the problem more complex than past approaches could handle.

Building on these challenges, the project team at Pacific Northwest National Laboratory (PNNL) has advanced previous research to enhance the practical applicability of topology discovery methods for real-world distribution systems. In prior work, the authors developed a graph-theory-informed state-estimation-based model selection approach [6]. Building upon this foundation, the team further hardened the algorithm to ensure robustness and suitability for utility deployment. The enhanced approach was demonstrated in collaboration with Peninsula Light Company and Survalent, using real feeder models and sensor data to validate its performance under practical operating conditions. This paper presents the methodology, demonstration results, and key insights gained from this utility-focused evaluation.

II. Topology Discovery Algorithm

The proposed methodology for topology identification is summarized in Fig. 1. The process is divided into two stages: a) graph-based candidate-topology enumeration and b) SE-based model selection (SEMS) on the candidate topologies. The first stage is carried out in two steps: i) model reduction and ii) case reduction. A brief overview of these steps is provided below; detailed formulations can be found in [6].

A. Model Reduction

In the first step of the topology identification process, a graph is constructed from a given planning model leveraging

979-8-3315-8138-1

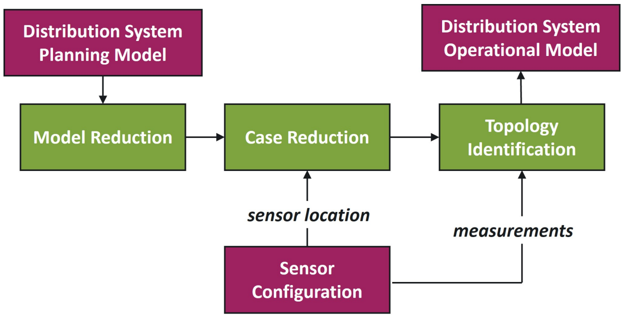

Fig. 1. Topology Discovery Algorithm.

equipment connectivity information. With the switch locations identified from the planning model, the graph is reduced such that only the switch edges are preserved. A connectivity matrix, obtained by summing powers of the adjacency matrix, is computed for each possible switch combination to identify the connectivity between the source and each node in the reduced-order graph. This information is stored for later use in the case reduction step.

B. Case Reduction

With node connectivity to the source established for each switch combination, measurements can be used to discard switch combinations. When a sensor or group of sensors within one node of the reduced-order network graph indicates that the node is energized, switch combinations without connectivity between the source and the node in question can be removed from consideration. Note that no assumptions are made regarding radial connectivity. Once the measurements from different sensors are obtained, case reduction is executed in parallel, invalid switch combinations are discarded, and the candidate topologies (Ω) are passed into model selection.

C. Model Selection

Once candidate topologies are identified, system states are calculated based on the given measurements and the network model. The objective function (or fitness function) denoted by O is calculated for each candidate topology, where the objective function is defined as the difference between the measurements and the system state. Mathematically,

$$O_w = \sum_{i \in B} \alpha_i \big(h_i(S) - z_i\big)^2 + \sum_{j \in V_c} \gamma_j \big(h_j(v) - z_j\big)^2, \quad \forall \omega \in \Omega \quad (1)$$

where B is the set of sensors measuring line flow; V_c is the set of sensors recording node voltages; S represents the line flow; and v represents nodal voltage. Since these measurements have different errors associated with them, corresponding weights (α_i, γ_j) are applied to bias the objective toward a specific measurement. First, a topology is extracted from the candidate topology list (Ω) using the original planning model. A system admittance matrix for the candidate topology is then constructed, leveraging the switching combinations corresponding to this topology. State estimation is solved for the candidate topology to determine the system states. The objective function calculates the sum of squared errors between system measurements and their respective variables as shown in (1). The candidate topologies are ranked according to their optimal objective value, and the one with the lowest objective value is identified as the optimal topology.

D. Metrics

The effectiveness of the proposed approach is thoroughly tested for different sets of unknown switches and varying sensor configurations. The metrics defined are capable of providing a measure of how far, on average, the identified topology is from the accurate topology if it is mis-detected.

1) Missed Switch Detection Ratio: Note that an incorrect estimation of one or more switches renders the given topology incorrectly identified. Thus, we define a metric, the missed switch detection ratio (MSDR), which quantifies the algorithm's performance in correctly estimating switch statuses, as shown in (2).

$$MSDR = \frac{S_C}{S \times M} \quad (2)$$

S_C is the number of switches identified incorrectly, S is the total number of switches, and M is the total number of trials.

2) Average Percentile Ranking: For a given topology, a percentile ranking (PR_i) of the accurate topology is calculated using the objective value at each trial. If $O = \{O_1, O_2, ..., O_N\}$ is the set of objective values of possible N topologies, the percentile rank of the correct topology with objective value O_i is given by:

$$PR_i = \left(1 - \frac{L}{N}\right) \times 100 \quad (3)$$

where L is the number of values in O that is lower than O_i and N is the total number of possible topologies. Once PR_i is obtained for each trial, an average percentile ranking is calculated using:

$$APR = \frac{\sum_{i=1}^{M} PR_i}{M} \quad (4)$$

3) Absolute Objective Difference: During a topology misdetection, a model with the smallest objective function value is not the correct topology. In such a case, there is a difference between the objective value of a correct topology and the topology with the least objective value. This metric quantifies the absolute difference between the residuals of the actual topology and the identified topology. Mathematically, AOD is given as,

$$AOD = |Min(O) - O_i| \quad (5)$$

where O_i is the objective value of the correct topology.

III. Modeling and Data Integration

The proposed workflow for modeling and data integration is illustrated in Fig. 2. The process began with extracting the distribution network model from Geographic Information System (GIS) data. The distribution network model used in this study was originally maintained in a proprietary format within a commercial distribution system simulation tool. This tool offers an export feature for OpenDSS, enabling the conversion of the network model into a set of DSS files.

979-8-3315-8138-1

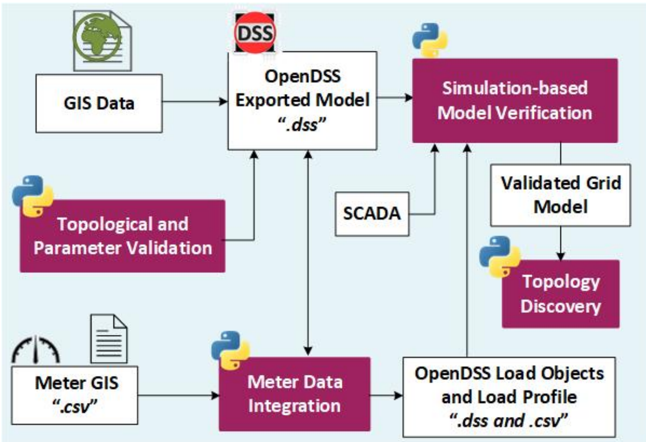

Fig. 2. Modeling and meter data integration.

A. Topology and Parameter Validation

This function thoroughly checks topology, base voltages, and component parameters to ensure the operational accuracy of the network representation. *Topology validation* involves detecting issues such as isolated components, phase inconsistencies, or disconnected buses. When missing connectivity is found, utility partners are consulted to update the model and ensure a complete network representation. *Nominal voltage validation* focuses on confirming that the primary and secondary voltages of transformers, as well as the nominal voltages assigned to other components—such as capacitors, loads, distributed energy resources, and voltage regulators—are correctly specified. Special attention is given to the control modes of regulators to ensure consistency with field operation. *Component parameter validation* includes verifying that critical physical and electrical parameters are reasonable and correctly assigned. This includes verifying that line impedance parameters, such as insulation values and geometric mean radius, transformer reactance values, and line lengths, are non-negative and non-zero where applicable, and that their units are consistent and accurate.

B. Meter Data Integration

SCADA data is used to drive the simulation for SCADA-connected loads or DERs, source voltage, and tap positions. The SCADA measurements are mapped to the corresponding equipment in the network model.

Similarly, AMI integration begins with the collection of time-series data from individual customer meters. Although this AMI data may not be available in real time, it can typically be exported in a structured format, such as a CSV file. To associate this meter data with the physical distribution network, PenLight meter GIS information is leveraged. This allows us to map each asset ID (meter ID) to its corresponding service transformer, enabling us to determine which transformer each customer meter is connected to. As multiple meters can be associated with a single service transformer, each meter is linked to a specific triplex node within the network model. This creates a direct connection between meter readings and precise locations in the network topology.

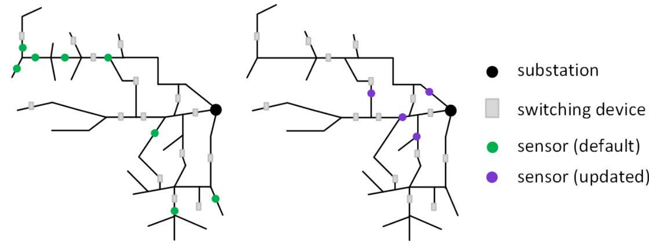

Fig. 3. Representative single line diagram of feeder with sensor location.

IV. Demonstration

A. Simulation Setup

Fig. 3 presents a one-line diagram that closely represents the feeder topology, including switching devices and sensor locations. While the number of switching devices and sensors is not exact, their placement and density are shown to approximately reflect those in the utility's model. In this work, we employed a deterministic estimation approach, where a load flow analysis is conducted to evaluate the theoretical values of line flows and node voltages that are consistent with the observed measurements. For the simulation purpose, a set of 20 different unknown switches was identified.

The AMI meter readings and SCADA measurements were aligned by selecting a timestamp at which most AMI data were available. Since AMI readings were recorded at 1-minute intervals, any missing values were filled by averaging the readings immediately before and after the target timestamp. For this demonstration, the timestamp 2025-06-25 17:30 was chosen, where SCADA data were available and approximately 95% of AMI readings were present. For the remaining 5%, interpolation using adjacent timestamps was applied to create a complete AMI dataset. If the operator intends to run this algorithm in real-time, forecasting will be necessary, as AMI data may not be available instantly. Alternatively, the operator can select a time window of a few minutes to an hour earlier, assuming the topology has not changed during that period, when the last AMI data was available. This approach allows the use of actual AMI data without relying on forecasting.

B. Simulation Results

In this section, we discuss the simulation results from two different studies. First, we identify the system's topology based on the measurements collected from SCADA at the chosen timestamp. Next, we create a hypothetical sensor configuration to quantify the tool's performance and compare it to the base sensor configuration.

1) Base Case: In this case, we assumed that around 20 switches had unknown statuses, representing those for which the operator might be uncertain about their actual state. With the given sensor configuration (telemetered reclosers in PenLight feeders and AMI readings), we were able to deduce seven switches from the list of unknown. This brings the total number of unknowns to 13. Now, out of 8,192 switching combinations, the case reduction stage provided only 256 valid combinations, as shown in Table I. The objective function value was computed for these 256 candidate topologies.

979-8-3315-8138-1

TABLE I
ALGORITHM CHARACTERISTICS FOR BASE CASE

Parameters	Value
Total Unknown Switches	20
Switches Deduced	7
Remaining Unknown Switches	$20 - 7 = 13$
Total Combinations	$2^{13} = 8192$
Total Valid Combinations	256

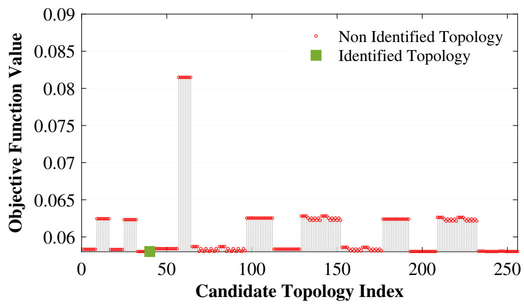

Fig. 4. Objective function plot for the base case.

Figure 4 shows the plot for the objective functions for all 256 valid combinations. The minimum objective function was observed for 39^{th} index, and that switching configuration represents the most likely topology. It is also observed that the objective function values for different candidate topologies are close to each other. This means that the identified topology may not always accurately represent the actual topology, as measurements are prone to errors. Since these measurements were obtained for a particular time, it was not practical to validate this output with the actual topology in the field, as the system's topology is not known or recorded for that time.

To test a known scenario, we created an operational topology in OpenDSS and generated noisy measurements using the base sensor configuration. With this, we would be able to compare the identified topology with the actual topology. Figure 5 shows such a case, where the identified topology and actual topology were different. As noted earlier, when measurements cannot clearly differentiate between topologies, incorrect identification may occur. Because errors are inherent in measurements and some topologies are very similar—meaning sensor readings are minimally affected by different switching configurations—misdetections are likely to occur. This highlights the importance of optimal sensor placement to enhance the accuracy of the topology discovery tool. The minimum number of measurements required to guarantee topological identifiability (assuming error-free measurements) is based on a spanning tree and a fundamental cycle [7]. The condition specifies that the measurement set must be chosen such that every possible spanning tree of the network differs from the others by at least one flow measurement, ensuring that each tree can be uniquely identified. Consequently, this implies that at least one flow measurement is needed along each fundamental cycle [3].

2) Updated Sensor Configuration: In this section, we present simulation results for the updated sensor placement, as shown in Fig. 3 (right). We placed sensors in various locations

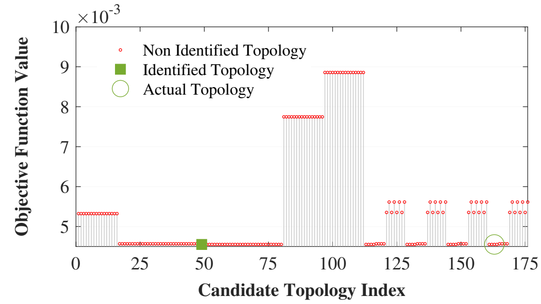

Fig. 5. Objective function value plot, where identified and actual topologies are different.

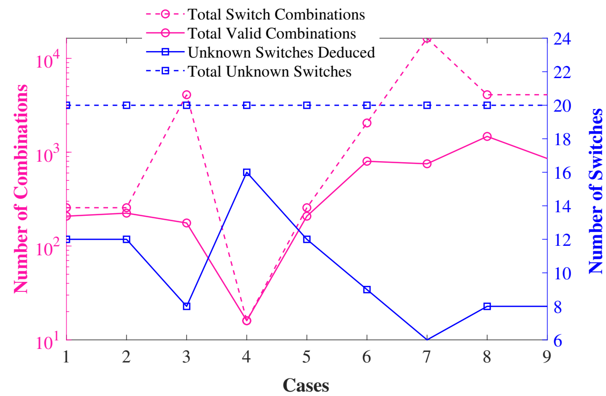

Fig. 6. Switching combination analyses for different cases.

to ensure that the configuration of candidate topologies would affect the sensor readings. It should be noted that we did not use an optimal sensor placement approach; instead, we applied a heuristic method to position sensors so that changes in topology would affect their readings. Flow readings were extracted from these sensor locations to create input for the tool, like in the previous case. Different cases were created, where each case represents a different set of unknown switches. The number of unknown switches is fixed to 20. Figure 6 shows the plot for different cases where the performance of the first stage of the algorithm is highlighted. In each case, the solid blue curve represents the number of unknown switches that were deduced during the first stage, while the solid red curve indicates the number of valid switch configurations retained for further analysis. For example, in Case 3, 8 unknown switches were deduced, and 12 unknown switches were considered for generating candidate topologies. Out of the 4,096 total possible combinations (illustrated by the dashed red curve), only 176 valid combinations were retained for the computation of objective functions (see solid red curve).

Fig. 7 shows the objective functions plot for all 176 valid combinations for case 3. It is observed that the identified topology and the actual topology were the same, thus indicating that the topology was accurately discovered.

On the other hand, it was observed that the identified topology and actual topology were different for case 6, as shown in Fig. 8. This indicates that some switches were identified incorrectly when compared to the actual topology. To investigate the cause of these discrepancies, a root cause analysis was performed. For the incorrectly identified switch,

979-8-3315-8138-1

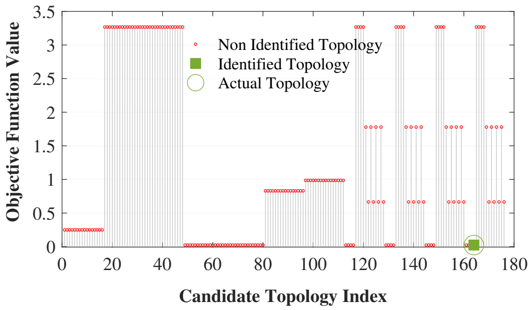

Fig. 7. Objective functions plot for case 3 in Fig. 6.

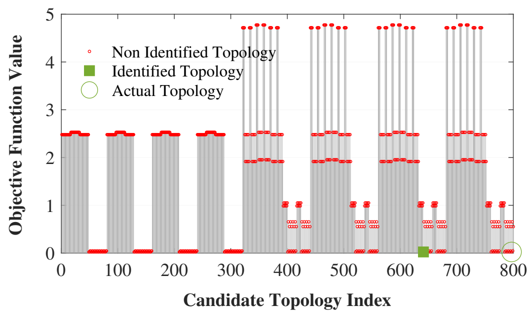

Fig. 8. Objective functions plot for case 6 in Fig. 6.

the observed flow was zero since no loads were connected downstream. As a result, the objective function was largely unaffected by whether the switch was open or closed. Consequently, the algorithm classified the switch as open, even though it was closed in the actual topology.

C. Performance Analysis

The three metrics described earlier were evaluated for different sensor configurations, as shown in Table II. The purpose of these metrics is to assess the performance of the topology discovery tool, particularly in scenarios where the correct topology is not explicitly identified. Even when the accurate topology is not determined, this performance analysis allows operators to examine the set of most likely topologies and understand the operational differences between the actual topology and these identified candidates.

For instance, the percentile ranking metric quantifies how many topologies exhibit an objective function value smaller than that of the actual topology. This statistical insight helps operators narrow down the list of most plausible topologies. Similarly, the MSDR metric represents the proportion of switches that were incorrectly identified. Finally, a lower AOD metric indicates that the identified topology is closely aligned with the actual topology in terms of objective function values. This similarity suggests that the identified topology has comparable operational characteristics to the actual topology and does not significantly affect grid states. As observed in Table II, the metrics improve with the updated sensor configuration. This improvement occurs because, in this case, the sensors are strategically positioned to ensure that the objective values of different candidate topologies differ more significantly. In the default configuration, the sensor locations were unable to distinguish among candidate topologies because their

TABLE II
PERFORMANCE COMPARISON FOR DEFAULT AND UPDATED SENSOR CONFIGURATIONS.

Metrics	Default Sensor Config.	Updated Sensor Config.
Percentile Ranking	61.41	10.21
MSDR	0.26	0.12
AOD	0.000298	$3.27e^{-7}$

readings showed minimal variation across configurations, and measurement errors further masked those small differences. In contrast, with the optimized sensor placement, the sensor readings vary with changes in topology, enabling the algorithm to more effectively identify the correct operational state and thus improve the overall performance metrics.

V. Conclusions

This paper presented a demonstration of a topology discovery algorithm developed and tested in collaboration with Survalent and Peninsula Light Company using real feeder models and sensor data. The results demonstrated the potential of advanced topology discovery tools to improve situational awareness in distribution system operations. The study underscored that sensor placement plays a pivotal role in determining algorithmic accuracy. However, many sensors currently installed in distribution systems - such as reclosers - are primarily deployed for protection or operational purposes and may not always provide the measurements required for accurate topology discovery.

To address this gap, the work emphasized the importance of conducting sensitivity analyses to guide cost-effective sensor deployment strategies that support multiple operational objectives while maintaining system observability. While challenges remain in handling sparse or imperfect field data, the demonstration provides a foundation for incremental advancements and field implementation. Ultimately, the work lays the groundwork for incorporating topology discovery capabilities into commercial platforms, enhancing grid reliability, visibility, and operational efficiency.

References

[1] V. Zamani and T. Nielsen, "How to create an accurate network model and dynamic state data for an advanced distribution management system (adms)," 2020.

[2] Z. Li, S. Jazebi, and F. De Leon, "Determination of the optimal switching frequency for distribution system reconfiguration," *IEEE Transactions on Power Delivery*, vol. 32, no. 4, pp. 2060–2069, 2016.

[3] A. Gandluru, S. Poudel, and A. Dubey, "Joint estimation of operational topology and outages for unbalanced power distribution systems," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 605–617, 2019.

[4] A. Srivastava, S. Chakrabarti, J. Soares, and S. N. Singh, "An optimization-based topology error detection method for power system state estimation," *Electric Power Systems Research*, vol. 209, p. 107914, 2022.

[5] H. S. Karimi and B. Natarajan, "Joint topology identification and state estimation in unobservable distribution grids," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5299–5309, 2021.

[6] S. Poudel, T. Ramachandran, A. Veeramany, C. Francis, and A. P. Reiman, "Topology identification using graph theory informed state estimation-based model selection for power distribution systems," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 3, pp. 3563–3573, 2023.

[7] R. A. Sevlian and R. Rajagopal, "Distribution system topology detection using consumer load and line flow measurements," *arXiv preprint arXiv:1503.07224*, 2015.

979-8-3315-8138-1