أكاديمية كاوست
KAUST ACADEMY

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Automated Vulnerability Discovery and Remediation in Power Grids using Semantic Automata Learning and GNNs

Meshal Alsulami, dr.Mohhammed Reda

meshalalsulme@gmail.com,

Abstract

We present SemanticGrid, a framework that bridges the gap between deep learning and formal verification for power grid stability. Modern AC grids are vulnerable to cascading failures; while Graph Neural Networks (GNNs) offer rapid contingency predictions, their "black-box" nature makes them unsuitable for mission-critical control rooms. To solve this, we trained a PyTorch Geometric GNN on physical simulations (ANDES) to act as a high-speed surrogate Membership Oracle. We then deployed the L^* automata learning algorithm to extract a formal, white-box Moore machine representing the IEEE 118-bus system. To ensure physical accuracy, a Counterexample-Guided Inductive Synthesis (CEGIS) loop verified Equivalence Queries against the ground-truth simulator. Our approach pruned 5,633 redundant topological transitions into a concise operational protocol. As a headline result, our $O(1)$ FSM protocol lookup achieves a 467,042x speedup over traditional Newton-Raphson N-k analysis. The model successfully isolated 1 critical vulnerability (the Betsy Layne transmission corridor, Line 131) and autonomously synthesized 2 verified remedial switching actions (Lines 36 and 38) to restore system voltage to a safe ≥ 0.95 p.u. baseline in fractions of a microsecond.

Date: August 19, 2026

Contents

1 Introduction **2**

2 Related Work **2**

3 Methodology **2**

3.1 Method Details 2

3.2 Evaluation Procedure 3

4 Results **3**

4.1 Performance and Scalability Benchmarks 3

4.2 Vulnerability Discovery and Remediation 3

5 Discussion **3**

6 Limitations and Future Work **3**

7 Conclusion **4**

1 Introduction

The stability of AC power grids relies on maintaining bus voltages within strict regulatory tolerances (typically $\pm 5\%$). When a transmission line trips due to a fault, the resulting redistribution of power flow can cause localized voltage drops, potentially triggering a cascading blackout.

To prevent this, operators rely on N-1 contingency analyses. However, modern grids are too large to simulate sequential N-k failures in real-time. For the IEEE 118-bus system, calculating N-3 contingencies using traditional Newton-Raphson solvers requires evaluating over 1.05 million scenarios, taking upwards of 4.4 hours—rendering it useless for split-second control room intervention.

While Graph Neural Networks (GNNs) solve this speed bottleneck, they operate as black boxes without formal guarantees. This project introduces a framework that uses a GNN purely as a high-speed proxy to accelerate formal Automata Learning. By extracting a human-readable Finite State Machine (FSM), we provide grid operators with instantaneous, mathematically verified recovery protocols for cascading failures.

2 Related Work

Recent advancements in smart grids have heavily leveraged Graph Neural Networks to predict power flow and identify vulnerabilities due to their spatial capabilities. However, these models lack the transparency required for critical infrastructure.

Conversely, Automata Learning (specifically the L^* algorithm) is widely used in software engineering to extract white-box formal models from complex systems. Our work bridges the gap by treating physical line trips as an input alphabet and grid voltage health as the system language, applying software verification techniques directly to physical transmission topologies.

3 Methodology

3.1 Method Details

To prevent state explosion, we mapped the absolute minimum voltage of the grid to four semantic operational states based on standard engineering thresholds:

- **NORMAL**: ≥ 0.95 p.u. (Safe operating band)
- **ALERT**: $0.90 \leq V < 0.95$ p.u. (System stressed)
- **EMERGENCY**: $0.85 \leq V < 0.90$ p.u. (Active cascade threat)
- **COLLAPSE**: < 0.85 p.u. (Point of no return)

GNN as a Membership Oracle: The L^* algorithm requires asking the system thousands of "Membership Queries" (e.g., "What is the grid state if Lines 129 and 131 trip?"). Querying a physical simulator for every hypothesis is computationally infeasible. Instead, we trained a GNN to act as a high-speed surrogate model, answering L^* queries in 0.0123 milliseconds.

Equivalence Queries and CEGIS Verification: To ensure the L^* algorithm converges correctly, it must execute Equivalence Queries to verify the proposed FSM. Because GNNs can hallucinate, we wrapped the equivalence phase in a Counterexample-Guided Inductive Synthesis (CEGIS) loop. The loop verifies the FSM's critical paths against the ground-truth ANDES simulator. If the GNN hallucinated, the CEGIS loop retrains the GNN on the true physical physics and restarts the extraction.

3.2 Evaluation Procedure

- **Tools Used**: ANDES (ground-truth physical simulation), PyTorch Geometric (GNN Oracle), and AALpy (L^* extraction).
- **Benchmarking**: Execution times were compared between traditional Newton-Raphson solvers, GNN inference, and FSM dictionary lookups for 186 transmission lines.

4 Results

Following CEGIS verification (which confirmed zero semantic hallucinations), the L^* algorithm successfully modeled the IEEE 118-bus system as a Moore machine.

4.1 Performance and Scalability Benchmarks

As shown in Table 1, traditional N-k analysis suffers from severe combinatorial explosion. Extracting the system into an FSM reduced complex N-k failure analysis from a 4.4-hour computational bottleneck to an instantaneous $O(1)$ dictionary lookup.

Table 1 Computational Benchmarks for N-k Contingency Analysis (IEEE 118-Bus System).

Methodology	Time per Eval	Speedup vs. Baseline
Newton-Raphson Simulation (Baseline)	15.000 ms	1x
GNN Surrogate Inference	0.0123 ms	1,224x
Semantic FSM Protocol Lookup (Ours)	0.000032 ms	**467,042x**

4.2 Vulnerability Discovery and Remediation

By pruning 5,633 redundant lateral edges, the FSM isolated a critical structural flaw:

- **The Vulnerability:** Tripping **Line 131** (the physical corridor connecting Betsy Layne to Beaver Creek) instantly drops the 118-bus grid from NORMAL into a systemic ALERT state.
- **Automated Remediation:** The FSM autonomously synthesized a verified recovery path. To halt the cascade from Line 131, the operator must execute a controlled trip of **Line 36** or **Line 38** (the Sorenson tie-lines), which reroutes bulk power flow and restores the grid to NORMAL.

5 Discussion

The results highlight a fundamental advantage of formal extraction. Traditional simulation approaches require calculating over 1.05 million scenarios to evaluate N-3 contingencies. By extracting the physical realities of the grid into a semantic Moore machine, we bypass the physical solvers entirely during real-time operations. The FSM does not merely flag Line 131 as dangerous; it maps the exact physical firewall (Lines 36/38) needed to save the system, providing operators with a deterministic, verified, and instantaneous control playbook.

6 Limitations and Future Work

- **Limitation:** Alphabet scalability. Using all 186 individual transmission lines as the input alphabet generated massive, complex topological trees that required significant post-processing to become human-readable.

- **Concrete Next Step:** We propose implementing Macro-Action Abstraction (Zoning). By grouping physical lines into a smaller alphabet of 5-10 logical regional actions (e.g., "Trip Inter-Area Tie Line"), the L^* algorithm will extract much more compact protocol diagrams naturally.
- **Future Direction:** Integrating the extraction pipeline directly with live SCADA/PMU data to dynamically update the vulnerability FSM as grid demand shifts.

7 Conclusion

This project successfully solved the black-box limitation of deep learning in power grid stability. By utilizing a GNN as a high-speed Membership Oracle and enforcing physical truth via a CEGIS loop, we autonomously extracted the operational protocol of the IEEE 118-bus system. The framework proved highly effective, achieving a 467,042x speedup over traditional solvers through $O(1)$ FSM lookups. The system isolated the Betsy Layne transmission corridor (Line 131) as a critical vulnerability and synthesized the Sorenson-tie remedial switching actions (Lines 36 and 38) required to halt cascading failures, providing a vital step toward verifiable, AI-assisted control rooms.