

# REAL-TIME RISKY FAULT-CHAIN SEARCH USING TIME-VARYING GRAPH RNNs

**Anmol Dwivedi**  
Rensselaer Polytechnic Institute

**Ali Tajer**  
Rensselaer Polytechnic Institute

## ABSTRACT

This paper introduces a data-driven graphical framework for the real-time search of risky cascading fault chains (FCs) in power-grids, crucial for enhancing grid resiliency in the face of climate change. As extreme weather events driven by climate change increase, identifying risky FCs becomes crucial for mitigating cascading failures and ensuring grid stability. However, the complexity of the spatio-temporal dependencies among grid components and the exponential growth of the search space with system size pose significant challenges to modeling and risky FC search. To tackle this, we model the search process as a partially observable Markov decision process (POMDP), which is subsequently solved via a time-varying graph recurrent neural network (GRNN). This approach captures the spatial and temporal structure induced by the system’s topology and dynamics, while efficiently summarizing the system’s history in the GRNN’s latent space, enabling scalable and effective identification of risky FCs.

## 1 INTRODUCTION

Large-scale disruptions in power-grids often stem from small, unnoticed anomalies that gradually stress the system. Past blackouts show that system operators often miss these signals, allowing failures to cascade into widespread disruptions [1]. Thus, forming real-time situational awareness is essential to preventing such failures and ensuring a resilient grid operation. This work enhances grid resiliency by dynamically predicting the risky fault chains the power system faces, leading to a more secure and reliable grid. A resilient grid is vital for scaling renewable energy integration, facilitating the transition to a low-carbon future, and playing a key role in mitigating climate change.

A fault chain (FC) [2] is a sequence of consecutive component outages that capture the temporal evolution of cascading failures in power-grids. Since a FC captures the high-impact, low-probability nature of cascades, timely identification of the riskiest FCs is crucial for predicting [3, 4, 5, 6] and mitigating cascading failures [7, 8, 9, 10]. Besides, FC search algorithms are widely used for risk assessment [11, 12, 13] and identifying vulnerable grid components [14, 15], among other applications [16]. Predicting FCs and assessing their risks involves (i) enumerating all possible failure scenarios up to a time horizon, (ii) evaluating the disruption (e.g., load loss) caused by each, and (iii) evaluating the likelihood of each scenario. However, accomplishing these three tasks face the following two key challenges: first, the scenario space grows exponentially with the system size and time horizon, making exhaustively listing all FCs and assessing their risk computationally infeasible, and second, the system’s dynamic nature necessitates constantly updating the scenario space.

There are two main approaches for the design of risky FC search algorithms: model-based and data-driven. Model-based methods qualitatively model, analyze and simulate large-scale cascading outage processes by developing simulation models that capture the system physics. For instance, detailed failure models such as the OPA [17], improved OPA [18], random chemistry [19] and others [20, 21, 22, 23, 24] precisely model the AC power-flow and dispatch constraints of the grid components while simulating cascades. Despite their detailed modeling and effectiveness, these models are computationally intensive, a major impediment to their adoption for *real-time* implementation. The next category of studies develop hybrid high-level statistical models to facilitate faster computation and quick inference such as the CASCADE model [25], the branching process model [26], the interaction model [27], and the influence graph models [4, 3]. While these approaches are quick in revealing important quantitative and interpretable properties of cascades, such models fail to accommodate the *time-varying* grid component interactions at different stages of a cascade.

Conversely, machine learning (ML) approaches have been proposed to enhance the efficiency of risky FC search. Broadly, these models assess the vulnerability of each power-grid component using simulated operational data via data-driven algorithms. For instance, the study in [28] formulates the search for risky FCs in a Markov decision process (MDP) environment and employs reinforcement learning (RL) algorithms to find risky FCs. The investigation in [29] employs deep neural networks to obtain critical states during cascading outages, and [30] employs convolutional neural networks (CNNs) for faster contingency screening. Authors in [15] propose a transition-extension approach that builds upon [28] to make it amenable to real-time implementation facilitated by exploiting the similarity between adjacent power-flow snapshots and finally, authors in [31] propose to employ graph-CNNs that leverage the grid’s topology to identify cascading failure paths.

Despite the effectiveness of the aforementioned ML models, these models broadly face the following challenges: (i) The models fail to capture the concurrent spatio-temporal dependencies across time among the grid components under dynamically *changing* topologies; (ii) the cascading outage process is assumed to follow Markov property. While this simplification can render reasonable approximations during the earlier stages of a cascade, the later stages of a cascade process typically exhibit temporal dependencies beyond the previous stage, making the Markovian assumptions inadequate; (iii) these models become prohibitive even for moderate grid sizes due to combinatorial growth in either the *computational* or *storage* requirements with the number of grid components.

**Contribution:** We propose a data-driven graphical framework for efficiently identifying risky cascading FCs to address the aforementioned challenges associated with the ML approaches. The proposed framework designs a graph recurrent neural network (GRNN) to circumvent the computational complexities of the real-time search of FCs. The search process is formalized as a partially observable Markov decision process (POMDP), which is subsequently solved via a time-varying GRNN that judiciously accounts for the inherent temporal and spatial structures of the data generated by the system. The key features of the model include (i) leveraging the spatial structure of the data induced by the system topology, (ii) leveraging the temporal structure of data induced by system dynamics, and (iii) efficiently summarizing the system’s history in the latent space of the GRNN, rendering the modeling assumptions realistic and the approach amenable to real-time implementation. The data and code required to reproduce our results is publicly available

## 2 PROBLEM FORMULATION

**Fault Chain Model:** Consider the topology of a generic outage-free system that precedes a FC given by  $\mathcal{G}_0$ , and the associated system state by  $\mathbf{X}_0 \in \mathbb{R}^{N \times F}$  where  $F$  and  $N$  denotes the number of system state parameters and buses (graph vertices), respectively. We denote the set of all components in the system by  $\mathcal{U}$ . We specify a generic FC that the system might be facing as a *sequence* of consecutive component outages consisting of at most  $P$  stages, where  $P$  can be selected based on the horizon of interest for risk assessment. In each stage  $i \in [P] \triangleq [1, \dots, P]$ , the set  $\mathcal{U}_i$  represents the components that fail during stage  $i \in [P]$ . Note that the set  $\mathcal{U}_i \subseteq \mathcal{U} \setminus \{\cup_{j=1}^{i-1} \mathcal{U}_j\}$  could consist of more than one component failure in any stage  $i$ , i.e.,  $|\mathcal{U}_i| \geq 1$ . Consequently, we denote the sequence of components that fail during the  $P$  stages by a FC sequence  $\mathcal{V} \triangleq \langle \mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_P \rangle$  and denote the healthy components before failure(s) in stage  $i$  of the FC sequence by  $\ell_i \in \mathcal{U} \setminus \{\cup_{j=1}^{i-1} \mathcal{U}_j\}$ .

**Quantifying Risk:** Due to component outages in each stage  $i$ , the system’s topology alters to  $\mathcal{G}_i \triangleq (V_i, E_i)$  with an associated adjacency matrix  $\mathbf{B}_i$  and an underlying system state  $\mathbf{X}_i$ . Consecutive failures lead to compounding stress on the remaining grid components, which need to ensure minimal load shedding in the network. Nevertheless, when the failures are severe, they lead to load losses. We denote the load loss (LL) imposed by the component failures in stage  $i$  by  $\text{LL}(\mathcal{U}_i) \triangleq \text{load}(\mathcal{G}_{i-1}) - \text{load}(\mathcal{G}_i)$ , where  $\text{load}(\mathcal{G}_i)$  is the *total* load (in MWs) when the system’s state in stage  $i$  is associated with the graph topology  $\mathcal{G}_i$ . Accordingly, we define the total load loss (TLL) imposed by any FC sequence  $\mathcal{V}$  by  $\text{TLL}(\mathcal{V}) \triangleq \sum_{i=1}^P \text{LL}(\mathcal{U}_i)$ .

**Maximizing Accumulated TLL:** Due to the re-distribution of power across grid components after each stage  $i$  of the FC, some FC sequences particularly lead to substantial risks, and owing to the continuously time-varying system’s state  $\mathbf{X}_0$ , different loading and topological conditions face different risks. Our objective is to identify  $S$  number of FC sequences, each consisting of  $P$  stages, that impose the *largest* TLL associated with any given initial observation  $\mathbf{O}_0 \triangleq (\mathcal{G}_0, \mathbf{X}_0)$ . To

formalize this, we define  $\mathcal{F}$  as the set of all possible FC sequences  $\mathcal{V}$  with a target horizon of  $P$ , and our objective is to identify  $S$  members of  $\mathcal{F}$  with the largest associated risk (load losses). We denote these  $S$  members by  $\{\mathcal{V}_1^*, \dots, \mathcal{V}_S^*\}_{\mathbf{O}_0}$ . Identifying the sets of interest can be formally cast as solving

$$\mathcal{P} : \quad \{\mathcal{V}_1^*, \dots, \mathcal{V}_S^*\}_{\mathbf{O}_0} \triangleq \arg \max_{\{\mathcal{V}_1, \dots, \mathcal{V}_S\} : \mathcal{V}_i \in \mathcal{F}} \sum_{i=1}^S \text{TLL}(\mathcal{V}_i). \quad (1)$$

Without loss of generality, we assume that the TLLs of the set of sequences  $\{\mathcal{V}_1^*, \dots, \mathcal{V}_S^*\}_{\mathbf{O}_0}$  are in the descending order, i.e.,  $\text{TLL}(\mathcal{V}_1^*) \geq \text{TLL}(\mathcal{V}_2^*) \geq \dots \geq \text{TLL}(\mathcal{V}_S^*)$ .

Solving  $\mathcal{P}$  faces computational challenges since the cardinality of  $\mathcal{F}$  grows exponentially with the number of components  $|\mathcal{U}|$  and horizon  $P$ . To address this, we design an agent-based learning algorithm that *sequentially* constructs the FCs  $\{\mathcal{V}_1^*, \dots, \mathcal{V}_S^*\}_{\mathbf{O}_0}$ . The agent starts by constructing  $\mathcal{V}_1^* \triangleq \langle \mathcal{U}_{1,1}, \dots, \mathcal{U}_{1,P} \rangle$  where it sequentially identifies the sets  $\{\mathcal{U}_{1,i}\}$  in each stage  $i$  (via our proposed Algorithm 1) by admitting the observation  $\mathbf{O}_0$  as its baseline input. The risk associated with each candidate set  $\mathcal{U}_{1,i}$  has two, possibly opposing, impacts. The first pertains to the immediate LL due to component failures in  $\mathcal{U}_{1,i}$ , and the second captures the LL associated with the future possible failures driven by the failures in  $\mathcal{U}_{1,i}$ . Hence, identifying the sets  $\mathcal{U}_{1,i}$  involves look-ahead decision-making and cannot be carried out greedily based on only the immediate LLs.

**Risky FC Sequential Search as a POMDP:** To control the computational complexity of the search process, each set  $\mathcal{U}_{1,i}$  is determined from the observation  $\mathbf{O}_i = (\mathcal{G}_i, \mathbf{X}_i)$  in the current stage  $i$  only. Since the observation  $\mathbf{O}_i$  in each stage provides only *partial* information for decision making despite the LL at stage  $i$  depending on *all* the past  $i-1$  stages and the set of components removed in those stages, we formalize the agent decision process by a partially observed Markov decision process (POMDP) characterized by the tuple  $(\mathcal{S}, \mathcal{A}, \mathbb{P}, \mathcal{R}, \mathcal{O}, \mathbb{Z}, \gamma)$ . Detailed information about the POMDP modeling techniques employed is provided in Appendix A.1. After the agent identifies  $\mathcal{U}_{1,i}$ , the agent removes all the components in this set to determine the updated observation  $\mathbf{O}_{i+1} = (\mathcal{G}_{i+1}, \mathbf{X}_{i+1})$  via physical simulation models, using which  $\mathcal{U}_{1,i+1}$  is determined. This process continues recursively for a total of  $P$  stages, at the end of which the set  $\mathcal{V}_1^*$  is constructed and the algorithm repeats this process  $S$  times to identify  $S$  FC sequences of interest. The policy  $\pi^*$  to finding the *riskiest* FC sequence  $\mathcal{V}_1^*$  starting from a baseline state  $\mathbf{S}_0$  (2) can be formally cast as solving  $\pi^* \triangleq \arg \max_{\pi} \mathbb{E}[V_{\pi}(\mathbf{S}_0)]$ .

### 3 GRAPHICAL RISKY FAULT-CHAIN SEARCH FRAMEWORK

**Motivation:** Model-free off-policy RL algorithms [32] with function approximation, such as deep  $Q$ -learning [33], are effective at finding good policies for high-dimensional state spaces  $\mathcal{S}$  without requiring access to the transition probability kernel  $\mathbb{P}$ . However, deep  $Q$ -learning is ineffective for solving (1) for two reasons: (i) it typically trains a policy to find the riskiest FC  $\mathcal{V}_1^*$  given a starting state  $\mathbf{S}_0$ , whereas we need the  $S$  riskiest FCs, and (ii) the partial observations  $\mathbf{O}_i$  at each stage  $i$  do not reflect the underlying POMDP state  $\mathbf{S}_i$  (2), leading to  $Q(\mathbf{O}_i, a_i) \neq Q(\mathbf{S}_i, a_i)$ .

**GRQN Architecture:** To exploit the underlying structure of each POMDP state  $\mathbf{S}_i$  (2), we leverage time-varying graph convolutional filters (7) to model the strong correlation induced by the meshed topology of power transmission networks as a latent state  $\mathbf{Z}_i \in \mathbb{R}^{N \times H}$ . We then model the evolution of this latent state via recurrence (9) to account for observational dependencies across stages in the FC sequence  $\mathcal{V}$ . Building on the approach of [34], we **estimate**  $\mathbf{Y}_i \in \mathbb{R}^{N \times G}$  (10) from  $\mathbf{Z}_i$  (9) at each stage  $i \in [P]$ , and use it to **predict**  $Q(\mathbf{Y}_i, a_i)$  as a proxy for the  $Q$ -values of each POMDP state-action pair  $Q(\mathbf{S}_i, a_i)$ , assembling into a graph recurrent  $Q$ -network (GRQN) architecture (more details in Appendix A.3.1).

**Graph Recurrent  $Q$ -learning for Risky FC Search:** To identify the  $S$  FC sequences with the highest TLL for a given initial observation  $\mathbf{O}_0$ , it is insufficient to merely **predict**  $Q(\mathbf{Y}_i, a_i)$ ; we must also **guide the search** for subsequent FCs with the next highest TLL by leveraging  $Q(\mathbf{Y}_i, a_i)$ . We propose a graph recurrent  $Q$ -learning Algorithm 1 to sequentially discover the  $S$  FC sequences of interest  $\{\mathcal{V}_1, \dots, \mathcal{V}_S\}_{\mathbf{O}_0}$  while concurrently updating the parameters of the GRQN via 12. To avoid repeatedly discovering the same FC sequences, we make two key modifications: first, we alter how “experience” is collected in the sequential experience buffer (Appendix A.3.2) by modifying the agent’s decision process, (13) and (14), to track how often each grid component  $\ell_i \in \mathcal{A}$  is removed from each POMDP state  $\mathbf{S}_i$  until the current search iteration; and second, instead of re-setting the

Algorithm	Range for Accumulated TLL $\sum_{i=1}^S \text{TLL}(\mathcal{V}_i)$ (in GWs)	Range for Regret( $S$ ) (in GWs)
Algorithm 1 ( $\kappa = 3$ )	110.42 $\pm$ 37%	765.33 $\pm$ 5.3%
Algorithm 1 ( $\kappa = 2$ )	96.18 $\pm$ 38%	779.57 $\pm$ 4.7%
Algorithm 1 ( $\kappa = 1$ )	86.43 $\pm$ 35%	789.32 $\pm$ 3.87%
PFW + RL + TE [15]	60.63 $\pm$ 3.4%	815.12 $\pm$ 0.26%
PFW + RL [15]	57.18 $\pm$ 3.1%	818.57 $\pm$ 0.22%

Table 1: Performance comparison for the IEEE-39 bus system under unbounded computational budget.

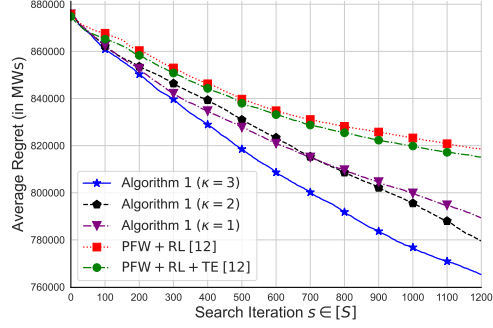


Figure 1: Regret( $s$ ) versus  $s$ .

latent states  $\mathbf{Z}_i$  to zero, we carry forward the previously learned latent state  $\mathbf{Z}_i$  of the GRNN after each newly discovered FC sequence  $\mathcal{V}_s$  during the parameter update (12) as the agent gains experience.

## 4 EXPERIMENTS

**Comparison under Unbounded Computational Budget:** We compare the accumulated TLL from the predicted FCs  $\{\mathcal{V}_1, \dots, \mathcal{V}_S\}_{\mathcal{O}_0}$  in  $S = 1200$  search iterations and benchmark the accuracy of Algorithm 1 against **ground-truth** by generating optimal FCs for the IEEE 39- and 118-bus systems. We calculate **regret** that quantifies the gap between the accumulated TLL from the **optimal** FCs  $\{\mathcal{V}_1^*, \dots, \mathcal{V}_S^*\}_{\mathcal{O}_0}$  and the predicted FCs. Until iteration  $s \in [S]$ , we define  $\text{Regret}(s) \triangleq \sum_{i=1}^S \text{TLL}(\mathcal{V}_i^*) - \sum_{i=1}^s \text{TLL}(\mathcal{V}_i)$ . Lower regret indicates a higher accuracy. Detailed descriptions of the algorithm, parameters, and baselines are in Appendix A.3, A.4 and A.5. Alongside Algorithm 1, we run  $Q$ -learning updates outlined in [15] for both PFW+RL and PFW+RL+TE baseline agents. Table 1 compares the accumulated TLL and regret (mean  $\pm$  SD) for various agents with **decreasing** computational complexity as we move down the table. Algorithm 1 consistently outperforms the baseline approaches, with  $\kappa = 3$  **reducing** average regret by 6.2% compared to the best baseline (PFW + RL + TE) and achieving almost **double** the accumulated TLL. We also observe that a greater  $\kappa$  results in **larger** accumulated TLL and **lower** regret on average since the weights of the behavior GRQN are updated more frequently (Algorithm 1), leading to better  $Q$ -value predictions for each POMDP state  $\mathbf{S}_i$ . For e.g., the average regret of Algorithm 1 with  $\kappa = 3$  is 765.3 GW, 3.04% lower than when  $\kappa = 1$ . Figure 1 further shows regret as a function of search iterations  $s \in [S]$ .

**Comparison under Bounded Computational Budget:** We next evaluate the performance of the agents within a strict **5-minute** computational time budget, simulating real-time implementation since the complexity for FC search should be within the dispatch cycle. Table 2 summarizes the results and three key observations emerge: (i) With a 5-minute budget, Algorithm 1 with  $\kappa = 3$  discovers **fewer** FC sequences compared to other algorithms. This is expected, as a higher  $\kappa$  requires more computation per iteration (Algorithm 1), reducing the total number of iterations and hence, discovered FC sequences  $S$ ; (ii) Algorithm 1 with  $\kappa = 2$  finds the **most** accumulated TLL. Although PFW + RL and PFW + RL + TE discover more FC sequences (1611 and 1608, respectively), their accumulated TLL are lower than Algorithm 1 with  $\kappa = 2$ ; (iii) In terms of accuracy, Algorithm 1 with  $\kappa = 3$  outperforms other algorithms, despite finding fewer FC sequences (575). The frequent weight updates of the GRQN lead to more accurate  $Q$ -value predictions, better **optimizing performance per search iteration**.

**Discussion:** It is noteworthy that all agents were evaluated on a standard computer **without** GPUs. Leveraging GPUs will further accelerate the search process for Algorithm 1, as they are designed for matrix and vector operations, unlike the baseline approaches, which cannot be similarly accelerated. We also emphasize that the baseline approach [15] models each *permutation* of component outages as a unique MDP state, and stores the  $Q$ -values for a combinatorial number of resulting MDP state-action pairs in an extensive  $Q$ -table, rendering it not scalable. However, by judiciously leveraging the graphical structure of each POMDP state and appropriately modeling the dependencies across the various stages of the FC, we have bypassed the storage challenge with fewer modeling assumptions while, at the same time, achieving better performance. Similar trends for the IEEE 118-bus system is provided in Appendix A.6, confirming the trends observed for the IEEE 39-bus system.

Algorithm	Average No. of FC Sequences $S$ Discovered	Range for Accumulated TLL $\sum_{i=1}^S \text{TLL}(\mathcal{V}_i)$ (in GWs)	Range for Regret( $S$ ) (in GWs)
Algorithm 1 ( $\kappa = 3$ )	575	63.484 $\pm$ 36.7%	457.09 $\pm$ 5.9%
Algorithm 1 ( $\kappa = 2$ )	700	79.792 $\pm$ 39%	521.47 $\pm$ 7.5%
Algorithm 1 ( $\kappa = 1$ )	937	70.848 $\pm$ 33.3%	673.56 $\pm$ 4.76%
PFW + RL + TE [15]	1608	68.74 $\pm$ 3.4%	965.41 $\pm$ 1.61%
PFW + RL [15]	1611	65.35 $\pm$ 3.9%	969.78 $\pm$ 2.27%

Table 2: Performance comparison for a computational time of 5 minutes for the IEEE-39 bus system.

## 5 CONCLUSION

In this paper, we have considered the problem of real-time risky fault chain identification in a limited number of search trials. We have proposed a data-driven graphical framework that can dynamically predict the chains of risky faults a power system faces. First, the search for risky fault chains is modeled as a partially observed Markov decision process (POMDP). Then a graph recurrent  $Q$ -learning algorithm is designed to leverage the grid’s topology to discover new risky fault chains efficiently. Experimental results on IEEE standard systems, compared to baseline methods, demonstrate the effectiveness and efficiency of our approach.

## REFERENCES

- [1] August 14, 2003 blackout: NERC actions to prevent and mitigate the impacts of future cascading blackouts. [https://www.nerc.com/docs/docs/blackout/NERC\\_Final\\_Blackout\\_Report\\_07\\_13\\_04.pdf](https://www.nerc.com/docs/docs/blackout/NERC_Final_Blackout_Report_07_13_04.pdf), February 2014.
- [2] Ansi Wang, Yi Luo, Guangyu Tu, and Pei Liu. Vulnerability assessment scheme for power system transmission networks based on the fault chain theory. *IEEE Transactions on Power Systems*, 26(1):442–450, 2011.
- [3] Xinyu Wu, Dan Wu, and Eytan Modiano. Predicting failure cascades in large scale power systems via the influence model framework. *IEEE Transactions on Power Systems*, 36(5):4778–4790, 2021.
- [4] Paul. D. H. Hines, I. Dobson, and P. Rezaei. Cascading power outages propagate locally in an influence graph that is not the actual grid topology. *IEEE Transactions on Power Systems*, 32(2):958–967, 2017.
- [5] Sathwik Chadaga, Xinyu Wu, and Eytan Modiano. Power failure cascade prediction using graph neural networks. In *2023 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pages 1–7, 2023.
- [6] Shiuli Subhra Ghosh, Anmol Dwivedi, Ali Tajer, Kyongmin Yeo, and Wesley M. Gifford. Cascading failure prediction via causal inference. *IEEE Transactions on Power Systems*, pages 1–12, 2024.
- [7] Kai Zhou, Ian Dobson, Zhaoyu Wang, Alexander Roitershtein, and Arka P. Ghosh. A markovian influence graph formed from utility line outage data to mitigate large cascades. *IEEE Transactions on Power Systems*, 35(4):3224–3235, 2020.
- [8] Deunsol Yoon, Sunghoon Hong, Byung-Jun Lee, and Kee-Eung Kim. Winning the 12{rpn} challenge: Power grid management via semi-markov afterstate actor-critic. In *International Conference on Learning Representations*, 2021. URL <https://openreview.net/forum?id=LmUJqB1Cz8>.
- [9] Anmol Dwivedi, Santiago Paternain, and Ali Tajer. Blackout mitigation via physics-guided RL. *IEEE Transactions on Power Systems*, pages 1–12, 2024.
- [10] Anmol Dwivedi, Ali Tajer, Santiago Paternain, and Nurali Virani. RL for mitigating cascading failures: Targeted exploration via sensitivity factors. *arXiv:2411.18050*, 2024.

- [11] Pooya Rezaei, Paul D. H. Hines, and Margaret J. Eppstein. Estimating cascading failure risk with random chemistry. *IEEE Transactions on Power Systems*, 30(5):2726–2735, 2015.
- [12] Pierre Henneaux, Pierre-Etienne Labeau, Jean-Claude Maun, and Liisa Haarla. A two-level probabilistic risk assessment of cascading outages. *IEEE Transactions on Power Systems*, 31(3):2393–2403, 2016.
- [13] Rui Yao, Shaowei Huang, Kai Sun, Feng Liu, Xuemin Zhang, Shengwei Mei, Wei Wei, and Lijie Ding. Risk assessment of multi-timescale cascading outages based on Markovian tree search. *IEEE Transactions on Power Systems*, 32(4):2887–2900, 2017.
- [14] Xiaoguang Wei, Junbo Zhao, Tao Huang, and Ettore Bompard. A novel cascading faults graph based transmission network vulnerability assessment method. *IEEE Transactions on Power Systems*, 33(3):2995–3000, 2018.
- [15] Zhimei Zhang, Rui Yao, Shaowei Huang, Ying Chen, Shengwei Mei, and Kai Sun. An online search method for representative risky fault chains based on reinforcement learning and knowledge transfer. *IEEE Transactions on Power Systems*, 35(3):1856–1867, 2020.
- [16] Lu Liu, Hao Wu, Linzhi Li, Danfeng Shen, Feng Qian, and Junlei Liu. Cascading failure pattern identification in power systems based on sequential pattern mining. *IEEE Transactions on Power Systems*, 36(3):1856–1866, 2021. doi: 10.1109/TPWRS.2020.3028999.
- [17] I. Dobson, B.A. Carreras, V.E. Lynch, and D.E. Newman. An initial model for complex dynamics in electric power system blackouts. In *Proc. Annual Hawaii International Conference on System Sciences*, Maui, HI, Jan 2001.
- [18] Shengwei Mei, Fei He, Xuemin Zhang, Shengyu Wu, and Gang Wang. An improved OPA model and blackout risk assessment. *IEEE Transactions on Power Systems*, 24(2):814–823, 2009. doi: 10.1109/TPWRS.2009.2016521.
- [19] Margaret J. Eppstein and Paul D. H. Hines. A “random chemistry” algorithm for identifying collections of multiple contingencies that initiate cascading failure. *IEEE Transactions on Power Systems*, 27(3):1698–1705, 2012.
- [20] S. Soltan, D. Mazauric, and G. Zussman. Analysis of failures in power grids. *IEEE Transactions on Control of Network Systems*, 4(2):288–300, 2017.
- [21] Yafei Yang, Xiaohong Guan, and Qiaozhu Zhai. Fast grid security assessment with  $N - k$  contingencies. *IEEE Transactions on Power Systems*, 32(3):2193–2203, 2017.
- [22] Tao Ding, Cheng Li, Chao Yan, Fangxing Li, and Zhaohong Bie. A bilevel optimization model for risk assessment and contingency ranking in transmission system reliability evaluation. *IEEE Transactions on Power Systems*, 32(5):3803–3813, 2017.
- [23] H. Cetinay, S. Soltan, F. A. Kuipers, G. Zussman, and P. V. Miegheem. Comparing the effects of failures in power grids under the AC and DC power flow models. *IEEE Transactions on Network Science and Engineering*, 5(4):301–312, 2018.
- [24] Jinpeng Guo, Feng Liu, Jianhui Wang, Junhao Lin, and Shengwei Mei. Toward efficient cascading outage simulation and probability analysis in power systems. *IEEE Transactions on Power Systems*, 33(3):2370–2382, 2018.
- [25] I. Dobson, B.A. Carreras, and D.E. Newman. A probabilistic loading-dependent model of cascading failure and possible implications for blackouts. In *Proc. Annual Hawaii International Conference on System Sciences*, Big Island, HI, Jan 2003.
- [26] I. Dobson, B.A. Carreras, and D.E. Newman. A branching process approximation to cascading load-dependent system failure. In *Proc. Annual Hawaii International Conference on System Sciences*, Big Island, HI, Jan 2004.
- [27] Junjian Qi, Kai Sun, and Shengwei Mei. An interaction model for simulation and mitigation of cascading failures. *IEEE Transactions on Power Systems*, 30(2):804–819, 2015.

- [28] Jun Yan, Haibo He, Xiangnan Zhong, and Yufei Tang. Q-learning-based vulnerability analysis of smart grid against sequential topology attacks. *IEEE Transactions on Information Forensics and Security*, 12(1):200–210, 2017.
- [29] Fangxing Li and Yan Du. From AlphaGo to power system AI: What engineers can learn from solving the most complex board game. *IEEE Power and Energy Magazine*, 16(2):76–84, 2018.
- [30] Yan Du, Fangxing Li, Jiang Li, and Tongxin Zheng. Achieving 100x acceleration for  $N - 1$  contingency screening with uncertain scenarios using deep convolutional neural network. *IEEE Transactions on Power Systems*, 34(4):3303–3305, 2019.
- [31] Yuxiao Liu, Ning Zhang, Dan Wu, Audun Botterud, Rui Yao, and Chongqing Kang. Searching for critical power system cascading failures with graph convolutional network. *IEEE Transactions on Control of Network Systems*, 8(3):1304–1313, 2021.
- [32] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- [33] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.
- [34] Matthew Hausknecht and Peter Stone. Deep recurrent  $Q$ -learning for partially observable MDPs. In *Proc. AAAI Fall Symposium Series*, Arlington, VA, July 2015.
- [35] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009.
- [36] Luana Ruiz, Fernando Gama, and Alejandro Ribeiro. Gated graph recurrent neural networks. *IEEE Transactions on Signal Processing*, 68:6303–6318, 2020.
- [37] Danilo Mandic and Jonathon Chambers. *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*. Wiley, 2001.
- [38] Anmol Dwivedi and Ali Tajer. Scalable quickest line outage detection and localization via graph spectral analysis. *IEEE Transactions on Power Systems*, 37(1):590–602, 2022.
- [39] Elvin Isufi, Andreas Loukas, Andrea Simonetto, and Geert Leus. Filtering random graph processes over random time-varying graphs. *IEEE Transactions on Signal Processing*, 65(16):4406–4421, 2017. doi: 10.1109/TSP.2017.2706186.
- [40] Fernando Gama, Qingbiao Li, Ekaterina Tolstaya, Amanda Prorok, and Alejandro R Ribeiro. Synthesizing decentralized controllers with graph neural networks and imitation learning. *IEEE Transactions on Signal Processing*, 2022.
- [41] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proc. International Conference on Learning Representations*, San Diego, CA, May 2015.
- [42] Anmol Dwivedi and Ali Tajer. GRNN-based real-time fault chain prediction. *IEEE Transactions on Power Systems*, 39(1):934–946, 2024.

## A APPENDIX

### A.1 POMDP MODELING

**State Space  $\mathcal{S}$ :** We denote the partial observation that the agent uses at stage  $i$  to determine the system state at stage  $i + 1$  by  $\mathbf{O}_i \triangleq (\mathcal{G}_i, \mathbf{X}_i)$ . Accordingly, we define the sequence

$$\mathbf{S}_i \triangleq \langle \mathbf{O}_0, \dots, \mathbf{O}_i \rangle, \quad (2)$$

which we refer to as the POMDP state at stage  $i$ , and it characterizes the entire past sequence of observations that render  $\mathbf{O}_{i+1}$ . As stated earlier, at stage  $i$  only  $\mathbf{O}_i$  is known to the agent.

**Action Space  $\mathcal{A}$ :** At stage  $i$ , upon receiving the observation  $\mathbf{O}_i$ , the agent aims to choose a component from the set of *available* components to be removed in the next stage. To formalize this process, we define the agent’s action as its choice of the component of interest. We denote the action at stage  $i$  by  $a_i$ . Accordingly, we define the action space  $\mathcal{A}_i$  as the set of all remaining (healthy) components, i.e.,  $\mathcal{A}_i \triangleq \mathcal{U} \setminus \{\cup_{j=1}^{i-1} \mathcal{U}_j\}$ . It is important to note that to reflect the reality of FCs, in which failures occur component-by-component, we are interested in identifying only *one* component in each stage  $i \in [P]$ . Nevertheless, due to the physical constraints, removing one component can possibly cause outages in one or more other components in the *same* stage. Hence, we denote the set of all components to be removed in stage  $i$  by the set  $\mathcal{U}_i$ .

**Transition Kernel  $\mathbb{P}$ :** Once the agent takes an action  $a_i \in \mathcal{A}_i$  in stage  $i$ , the underlying POMDP state in the next stage is randomly drawn from a transition probability distribution  $\mathbb{P}$

$$\mathbf{S}_{i+1} \sim \mathbb{P}(\mathbf{S} \mid \mathbf{S}_i, a_i). \quad (3)$$

The probability distribution  $\mathbb{P}$  captures the randomness due to the power system dynamics, and it is determined by the generator re-dispatch strategy employed in each stage of the FC. Nevertheless, for a given generator re-dispatch strategy and a given initial state distribution  $\mathbf{S}_0 = \mathbf{O}_0$  for which the  $\mathbf{S}$  FC sequences of interest are to be determined, the transitions kernel is *deterministic*.

**Reward Dynamics  $\mathcal{R}$ :** To quantify the risk associated with taking action  $a_i$  in POMDP state  $\mathbf{S}_i$  when transitioning to POMDP state  $\mathbf{S}_{i+1}$ , we define an instant reward  $r_i$

$$r_i \triangleq r(\mathbf{S}_{i+1} \mid \mathbf{S}_i, a_i) \triangleq \text{load}(\mathcal{G}_i) - \text{load}(\mathcal{G}_{i+1}). \quad (4)$$

Hence, for any generic action selection strategy  $\pi$ , the aggregate reward collected by the agent starting from the baseline POMDP state  $\mathbf{S}_0$  can be characterized by a value function

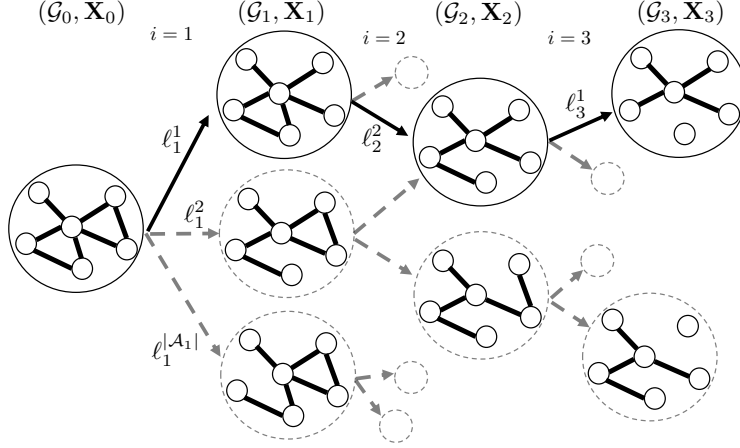
$$V_\pi(\mathbf{S}_0) \triangleq \sum_{i=0}^{P-1} \gamma^i \cdot r(\mathbf{S}_{i+1} \mid \mathbf{S}_i, \pi(\mathbf{O}_i)), \quad (5)$$

where the discount factor  $\gamma \in \mathbb{R}_+$  decides how much future rewards are favored over instant rewards, and  $\pi(\mathbf{O}_i)$  denotes the action selected by the agent given an observation  $\mathbf{O}_i$  in stage  $i \in [P]$ . Fig. 2 illustrates a search process where an agent constructs a FC sequence  $\mathcal{V}_s^* = \langle \ell_1^1, \ell_2^2, \ell_3^1 \rangle$  by leveraging the current system state  $(\mathcal{G}_i, \mathbf{X}_i)$  in each stage  $i \in [3]$ .

### A.2 TIME-VARYING GRNN MODEL

GRNNs are a family of GNN architectures [35] specialized for processing sequential graph structured data streams [36]. These architectures exploit the *local* connectivity structure of the underlying network topology  $\mathcal{G}_i$  to efficiently extract features from each bus by sequentially processing time-varying system states  $\mathbf{X}_i$  that evolve on a sequence of graphs  $\mathcal{G}_i$ . More broadly, these architectures generalize recurrent neural networks [37] to graphs. To lay the context for discussions, we first discuss time-varying graph convolutional neural networks (GCNNs), the components of which serve as an essential building block to motivate time-varying GRNNs.



Figure 2: An agent decision process rendering a FC  $\mathcal{V}_s^* = \langle \ell_1^1, \ell_2^2, \ell_3^1 \rangle$ .

**Time-Varying GCNNs:** Consider the  $i^{\text{th}}$  stage of a FC sequence  $\mathcal{V}_s$  where an agent receives an observation  $\mathbf{O}_i \triangleq (\mathcal{G}_i, \mathbf{X}_i)$  on graph  $\mathcal{G}_i$  associated with an adjacency matrix  $\mathbf{B}_i$ . Central to the development of GCNNs is the concept of a *graph-shift* operation that relates an input system state  $\mathbf{X}_i \in \mathbb{R}^{N \times F}$  to an output system state  $\mathbf{F}_i \in \mathbb{R}^{N \times F}$

$$\mathbf{F}_i \triangleq \mathbf{B}_i \cdot \mathbf{X}_i. \quad (6)$$

Clearly, the output system state  $\mathbf{F}_i$  is a locally shifted version of the input system state  $\mathbf{X}_i$  since each element  $[\mathbf{F}_i]_{u,f}$ , for any bus  $u \in V_i$  and parameter  $f$ , is a linear combination of input system states in its 1-hop neighborhood. Local operations, such as (6) capture the 1-hop structural information from  $\mathbf{X}_i$  by *estimating* another system state  $\mathbf{F}_i$  and are important because of the strong spatial correlation that exists between  $[\mathbf{X}_i]_{u,f}$  and its network neighborhood determined by  $\mathcal{G}_i$ . Alternative transformations such as that employed in [38] quantify the merits of estimating system states that capture the spatial structure of  $\mathbf{X}_i$ . In order to capture the structural information from a broader  $K$ -hop neighborhood instead, (6) can be readily extended by defining a time-varying *graph convolutional filter* function  $\mathbf{H} : \mathbb{R}^{N \times F} \rightarrow \mathbb{R}^{N \times H}$  that operates on an input system state  $\mathbf{X}_i$  to *estimate* an output system state  $\mathbf{H}(\mathcal{G}_i, \mathbf{X}_i; \mathcal{H})$

$$\mathbf{H}(\mathcal{G}_i, \mathbf{X}_i; \mathcal{H}) \triangleq \sum_{k=1}^K [\mathbf{B}_i^{k-1} \cdot \mathbf{X}_i] \cdot \mathbf{H}_k, \quad (7)$$

where  $H$  denotes the number of output features *estimated* on each bus and  $\mathcal{H} \triangleq \{\mathbf{H}_k \in \mathbb{R}^{F \times H} : k \in [K]\}$  denotes the set of filter coefficients parameterized by matrices  $\mathbf{H}_k$  *learned* from simulations where each coordinate of  $\mathbf{H}_k$  suitably weighs the aggregated system state obtained after  $k$  repeated 1-hop graph-shift operations performed on  $\mathbf{X}_i$ . Note that (7) belongs to a broad family of *graph-time filters* [39] that are polynomials in time-varying adjacency matrices  $\mathbf{B}_i$ . There exists many types of *graph-time filters* [40]. We employ (7) due to its simplicity. Nevertheless, (7) only captures simple linear dependencies within  $\mathbf{X}_i$ . To capture non-linear relationships within  $\mathbf{X}_i$ , time-varying GCNNs *compose* multiple layers of graph-time filters (7) and non-linearities such that the output system state of each GCNN layer is given by

$$\Phi(\mathcal{G}_i, \mathbf{X}_i; \mathcal{H}) \triangleq \sigma(\mathbf{H}(\mathcal{G}_i, \mathbf{X}_i; \mathcal{H})), \quad (8)$$

where  $\sigma : \mathbb{R} \rightarrow \mathbb{R}$  is commonly known as the activation function (applied element-wise) such that  $\Phi(\mathcal{G}_i, \mathbf{X}_i; \mathcal{H}) \in \mathbb{R}^{N \times H}$ .

**Time-Varying GRNNs:** GCNNs can only extract spatial features from each system state  $\mathbf{X}_i$  independently (8). For this reason, we add recurrency to our GCNN model (8) in order to capture the temporal observational dependencies across the various stages of a FC sequence to construct a GRNN. A time-varying GRNN extracts temporal features from an input sequence  $\langle \mathbf{O}_i : i \in [P] \rangle$



### A.3.2 SEQUENTIAL EXPERIENCE BUFFER

In order to deal with the issue of *catastrophic forgetting* [33], we employ a *sequential* experience buffer that stores FC sequences discovered during training of the GRQN from which batches of random sequences are sampled to facilitate the learning of parameters  $\theta$ . While there exists various ways to implement such a buffer, we employ an ordered list that stores all the visited transition tuples as a sequence  $\langle (\mathbf{O}_i, a_i, r_i, \mathbf{O}_{i+1}, \text{end}(\mathbf{O}_{i+1})) : i \in [P] \rangle$  where we have defined  $\text{end}(\mathbf{O}_{i+1})$  as a boolean value, if true, indicating that the observation  $\mathbf{O}_{i+1}$  is associated with a last stage of the risk assessment horizon  $P$ .

### A.3.3 TRAINING THE GRQN

For stability in the training of the GRQN, we employ the standard trick [33] of splitting the task of predicting and evaluating  $Q$ -values via two separate GRQNs, a target network  $\text{GRQN}(\cdot, \cdot | \theta^-)$  and a behavior network  $\text{GRQN}(\cdot, \cdot | \theta)$  each parameterized by a distinct set of parameters  $\theta^-$  and  $\theta$ , respectively. For every training iteration  $n$  of the graph recurrent  $Q$ -learning algorithm, the agent samples  $B$  random batches of FC sequences, each of type  $\langle (\mathbf{O}_j, a_j, r_j, \mathbf{O}_{j+1}, \text{end}(\mathbf{O}_{j+1})) : j \in [P] \rangle$ , from the sequential experience buffer on which the target GRQN is *unrolled* to estimate  $\mathbf{Y}_j$  (10) using which  $B$  batches of  $Q(\mathbf{Y}_j, \cdot | \theta^-)$  are predicted with respect to the target network. Subsequently, the agent computes a *look-ahead* target output for each batch where each target  $t_j \forall j \in [P]$  is given by

$$t_j = r_j + \gamma \cdot (1 - \text{end}(\mathbf{O}_{j+1})) \cdot \max_a Q(\mathbf{Y}_{j+1}, a | \theta^-). \quad (11)$$

Accordingly, the parameters of the behavior network is updated via gradient descent with respect to a quadratic loss

$$\theta_{n+1} = \theta_n - \alpha \cdot \nabla_{\theta} (t_j - Q(\mathbf{Y}_j, a_j | \theta))^2, \quad (12)$$

where  $\alpha$  denotes the learning rate and  $n$  denotes the current training iteration. The update (12) is preformed  $\kappa$  times for every action taken by the agent in any stage  $i \in [P]$  of the FC and additionally, serves as a means to control the *computational complexity* of our learning algorithm. In this paper, we employ the Adam optimizer [41] to perform the gradient update (12) and update the target network parameters  $\theta^- = \theta$  at the end of every FC sequence discovered.

### A.3.4 GRAPH RECURRENT LATENT SYSTEM STATE UPDATES

As the agent gains more experience and continues to store visited transition sequences of tuples in the experience buffer, the latent system state  $\mathbf{Z}_i$  of the GRNN may either be zeroed or carried forward after every newly discovered FC. Our experiments suggest that sequential updates where the latent system state  $\mathbf{Z}_i$  (9) is carried forward from the previous stages throughout the parameter update (12) leads to learning of better FC search strategies. Hence, we choose to carry forward the previously learned latent system state during the training of our GRQN.

### A.3.5 OUTLINE OF ALGORITHM 1

Algorithm 1 outlines the graph recurrent  $Q$ -learning algorithm used to *learn* the parameters  $\theta$  of the behaviour  $\text{GRQN}(\cdot, \cdot | \theta)$  to facilitate the discovery of the  $S$  number of FC sequences of interest. During the initial few iterations, since the sequential experience buffer is empty, we allow the agent to *explore* and fill the buffer with FC sequences for Explore number of iterations *offline* (more details in Section A.3.7). Subsequently, the real-time algorithm is initiated. Initially, the agent lacks any information about the cascading failure dynamics. Hence, it relies on the *prior knowledge* to construct  $\mathcal{U}_{j,i}$ , for any fault chain  $j$  in each stage  $i \in [P]$ , as the agent is unable to evaluate the LL associated with removing an arbitrary component  $\ell_i \in \mathcal{U}_j \setminus \{\cup_{k=1}^{i-1} \mathcal{U}_{j,k}\}$  in any stage of the FC  $\mathcal{V}_j$ . Gradually, Algorithm 1 learns to construct the sets  $\mathcal{U}_{j,i}$  by leveraging the learned latent graphical feature representations. These representations are obtained by optimizing the *look-ahead* target function (11) characterized by the behavior network  $\text{GRQN}(\cdot, \cdot | \theta)$  since the parameters  $\theta$  of this network determine the  $Q$ -values influencing the actions  $a_i \in \mathcal{A}_i$  taken by the agent. Therefore, when choosing actions  $a_i \in \mathcal{A}_i$ , the agent should make a trade-off between *exploration* and *exploitation* throughout the training of the GRQNs. Next, we discuss an *exploration-exploitation* search strategy that the agent employs to make the real-time search of FCs of interest more efficient.

**Algorithm 1** Graph Recurrent  $Q$ -learning (GRQN)

---

```

1: procedure GRAPH RECURRENT  $Q$ -LEARNING
2:   Initialize behaviour network with random  $\theta$  GRQN( $\cdot, \cdot | \theta$ )
3:   Initialize target network with weights  $\theta^- = \theta$  GRQN( $\cdot, \cdot | \theta^-$ )
4:   Initialize Buffer  $\leftarrow \langle \rangle$  from Appendix A.3.2
5:   Initialize  $\mathbf{Z}_0 \leftarrow \mathbf{0} \in \mathbb{R}^{N \times H}$ 
6:   for Episode  $s = 1, \dots, S$  do
7:     Episode  $\leftarrow \langle \rangle$ 
8:      $\mathcal{V}_s \leftarrow \langle \rangle$ 
9:     Reset power-flow according to initial state  $\mathbf{S}_0, \mathbf{O}_0$ 
10:    for Stage  $i = 1, \dots, P$  do
11:       $\_, \mathbf{Z}_i \leftarrow \text{GRQN}(\mathbf{O}_i, \mathbf{Z}_{i-1} | \theta)$ 
12:       $a_i \leftarrow \begin{cases} \text{explore via (13)} & \text{if } \text{rand}(0, 1) \leq \epsilon \\ \text{exploit via (14)} & \text{otherwise} \end{cases}$ 
13:      Take action  $a_i$  and determine  $\mathcal{U}_i$ 
14:       $\mathcal{V}_s \leftarrow \mathcal{V}_s \cup \mathcal{U}_i$ 
15:      Update power-flow and obtain  $\mathbf{O}_{i+1}$ 
16:      Calculate load loss  $r_i$  from (4)
17:      Episode  $\leftarrow \text{Episode} \cup (\mathbf{O}_i, a_i, r_i, \mathbf{O}_{i+1}, \text{end}(\mathbf{O}_{i+1}))$ 
18:      if  $s \geq \text{Explore}$  then
19:         $\text{count}(\mathbf{S}_i, a_i) \leftarrow \text{count}(\mathbf{S}_i, a_i) + 1$ 
20:        for training  $n = 1, \dots, \kappa$  do
21:          Sample  $B$  FC sequences from Buffer
22:           $t_j, \_ \leftarrow \text{GRQN}(\mathbf{O}_{j+1}, \mathbf{0} | \theta^-)$  from (11)
23:           $Q(\mathbf{Y}_j, a_j | \theta), \_ \leftarrow \text{GRQN}(\mathbf{O}_j, \mathbf{0} | \theta)$ 
24:          Calculate  $\nabla_{\theta}(t_j - Q(\mathbf{Y}_j, a_j | \theta))^2$ 
25:          Update  $\theta$  as in (12)
26:          Update  $\epsilon$  as in (15)
27:        end for
28:      end if
29:      Update availability of actions backwards
30:    end for
31:    Buffer  $\leftarrow \text{Buffer} \cup \text{Episode}$ 
32:     $\mathbf{Z}_0 \leftarrow \mathbf{Z}_P$ 
33:    if  $\text{TLL}(\mathcal{V}_s) \geq M$  then
34:      Store risky FC
35:    end if
36:     $\theta^- = \theta$ 
37:  end for
38: end procedure
39: Find the accumulated risk due to all the  $S$  FCs  $\{\mathcal{V}_1, \dots, \mathcal{V}_S\}$ .

```

---

## A.3.6 FAULT CHAIN SEARCH STRATEGY

We employ the standard  $\epsilon$ -greedy search strategy with an adaptive exploration schedule. Initially, the agent is compelled to take actions based on prior knowledge to find FCs with maximum expected TLL. Typically, since an outage of a component carrying higher power makes the remaining components vulnerable to overloading, a reasonable exploration strategy of the agent would be to remove components carrying maximum power-flow. Therefore, we follow a power-flow weighted (PFW) exploration strategy (also adopted in [15]) such that, in any stage  $i$  of the FC, the agent chooses the  $j^{\text{th}}$  available component  $\ell_i^j \in \mathcal{A}_i$  according to the rule

$$a_i = \arg \max_{\ell_i^j} \frac{\text{PF}(\ell_i^j) / \sqrt{\text{count}(\mathbf{S}_i, \ell_i^j) + 1}}{\sum_{k=1}^{|\mathcal{A}_i|} \text{PF}(\ell_i^k) / \sqrt{\text{count}(\mathbf{S}_i, \ell_i^k) + 1}}, \quad (13)$$

with probability  $\epsilon$  where we denote  $\text{PF}(\ell_i^j)$  as the *absolute* value of the power flowing through component  $\ell_i^j$  and denote  $\text{count}(\mathbf{S}_i, \ell_i^j)$  as the number of times the component  $\ell_i^j$  was chosen when the agent was in POMDP state  $\mathbf{S}_i$  in the past. On the other hand, as the agent gains more experience, the agent should choose actions based on the  $Q$ -values learned via the behaviour network  $\text{GRQN}(\cdot, \cdot | \theta)$ . Accordingly, a strategy based on  $Q$ -values learned by the agent is designed. Specifically, actions are chosen proportional to the  $Q$ -values normalized by each POMDP state-action visit count to *avoid*

repetitions of FC sequences discovered earlier. Accordingly, the agent chooses action  $a_i$

$$a_i = \arg \max_{\ell_i^j} \frac{Q(\mathbf{Y}_i, \ell_i^j | \boldsymbol{\theta})}{\sqrt{\text{count}(\mathbf{S}_i, \ell_i^j) + 1}}, \quad (14)$$

with probability  $1 - \epsilon$ .

In order to balance the exploration-exploitation trade-off between (13) and (14) during the training of the GRQN, it is important to dynamically alter the probability  $\epsilon$  so that, with more experience, the agent chooses actions based on (14). Appropriately, we follow the exploration schedule given by

$$\epsilon = \max \left( \frac{\sum_{j=1}^{|\mathcal{A}_1|} \text{PF}(\ell_1^j) / \sqrt{\text{count}(\mathbf{S}_0, \ell_1^j) + 1}}{\sum_{k=1}^{|\mathcal{A}_1|} \text{PF}(\ell_1^k)}, \epsilon_0 \right), \quad (15)$$

where  $\epsilon_0$  ensures a minimum level of exploration.

### A.3.7 OFFLINE SEQUENTIAL BUFFER INITIALIZATION

To initiate the parameter update of the behavior GRQN via gradient descent (12), there must exist at least  $B$  FC sequences in the experience buffer. However, unlike in the case of (13), the buffer can be populated *offline* for any loading condition. Therefore, the agent can afford to take actions greedily with respect to components conducting maximum power and, accordingly, backtrack to update the availability of actions to avoid repeating FCs discovered previously. Hence, prior to the start of our real-time FC search Algorithm 1, we let the agent explore *offline* for Explore iterations where the agent, in any stage  $i$ , chooses actions according to the rule

$$a_i = \arg \max_{\ell_i^j} \frac{\text{PF}(\ell_i^j)}{\sum_{k=1}^{|\mathcal{A}_i|} \text{PF}(\ell_i^k)}, \quad (16)$$

with probability 1 to fill the sequential experience buffer. Note that this needs to be done only once offline for any loading condition. This is important since the quality of the sequences in the buffer greatly affects the efficiency of the search.

## A.4 IEEE-39 NEW ENGLAND TEST SYSTEM

This test system comprises of  $N = 39$  buses and  $|\mathcal{U}| = 46$  components, including 12 transformers and 34 lines. We consider a loading condition of  $0.55 \times \text{base\_load}$ , where  $\text{base\_load}$  denotes the standard load data for the New England test case in PYPOWER *after* generation-load balance to quantify the performance of our approach. These loading conditions were chosen since it is relatively difficult to discover FCs with large TLLs in a lightly loaded power system as there are fewer such FCs in comparison to the space of all FC sequences  $|\mathcal{F}|$ .

### A.4.1 PARAMETERS AND HYPER-PARAMETERS

The hyper-parameters of the GRQNs are chosen by performing hyper-parameter tuning. Accordingly, we choose  $H = G = 12$  latent and output number of features when computing both the latent system state (9) and the output system state (10). We use  $K = 3$  graph-shift operations for the graph-filter (7) that is used to compute (9) and (10). We use both  $\rho$  and  $\sigma$  as the hyperbolic tangent non-linearity  $\sigma = \rho = \tanh$  and the ReLU non-linearity for the fully-connected NN that approximates the  $Q$ -values. For other parameters, we choose  $F = 1$  since we employ voltage phase angles as the only input system state parameter  $f$ , choose  $\gamma = 0.99$  since large LLs mostly occur in the last few stages of the FC sequence, an  $\epsilon_0 = 0.01$  to ensure a minimum level of exploration during the FC search process, a batch size  $B = 32$ , Explore = 250, a risk assessment horizon  $P = 3$ , FC sequence iteration  $S = 1200$  (*excluding* the initial Explore iterations), learning rate  $\alpha = 0.005$ , and  $\kappa \in [3]$  that controls the frequency of the learning update (12) and also governs the computational complexity of the graph recurrent  $Q$ -learning algorithm.

To quantify the regret, we need to compute the TLL associated with the  $S$  most critical FC sequences (ground truth)  $\mathcal{V}_s^*$ ,  $\forall s \in [S]$ . This is carried out by generating *all possible* FC sequences (i.e., set  $\mathcal{F}$ ) with a target horizon of  $P = 3$  for the considered total load of  $0.55 \times \text{base\_load}$ . By leveraging the pre-computed set  $\mathcal{F}$ , we observe a total of 3738 risky FCs for the loading condition  $0.55 \times \text{base\_load}$  using our developed FC simulator.

## A.5 BASELINE AGENTS

To further illustrate the merits of the proposed graphical framework, we compare it with two state-of-the-art baseline approaches proposed in [15]. We label the first approach in [15] based on the ordinary  $Q$ -learning algorithm without prior knowledge as PFW + RL and label their best performing approach based on transition and extension of prior knowledge from other power system snapshots by PFW + RL + TE. To ensure a fair comparison, we employ the *same* exploration schedule for  $\epsilon$  discussed in Section A.3.6 with the same parameters and the same discount factor  $\gamma$  for all the approaches. Note that, in the PFW + RL + TE approach, we first run their proposed  $Q$ -learning based approach offline, for a loading condition of  $0.6 \times \text{base\_load}$  (bringing in the prior knowledge). This is run for  $S = 5000$  iterations to ensure the convergence of their  $Q$ -learning algorithm. Subsequently, we store its extensive  $Q$ -table to run its PFW + RL + TE approach in real-time for the considered loading condition of  $0.55 \times \text{base\_load}$ , signifying a transition from the power system snapshot loaded at  $0.6 \times \text{base\_load}$  and an extension to the current power system snapshot loaded at  $0.55 \times \text{base\_load}$ . Note that, when performing comparisons, we set the parameters and hyper-parameters associated with Algorithm 1 the same as that described in Section A.4.1.

## A.6 IEEE-118 TEST SYSTEM

This test system consists of  $N = 118$  buses and  $|\mathcal{U}| = 179$  components. We consider a loading condition of  $0.6 \times \text{base\_load}$ , where  $\text{base\_load}$  denotes the standard load data for the IEEE-118 test case in PYPOWER after generation-load balance to quantify the performance of our approach.

### A.6.1 PARAMETERS AND HYPER-PARAMETERS

We choose  $H = G = 48$  latent and output number of features when computing both the latent system state (9) and the output system state (10). We use  $K = 3$  graph-shift operations for the graph-filter (7), use both  $\rho$  and  $\sigma$  as the hyperbolic tangent non-linearity  $\sigma = \rho = \tanh$  and the ReLU non-linearity for the fully-connected NN to approximate  $Q$ -values. For other parameters, we choose  $F = 1$  since we employ voltage phase angles as the only input system state parameter  $f$ , choose  $\gamma = 0.99$ ,  $\epsilon_0 = 0.01$ , a batch size  $B = 32$ , Explore = 250, a risk assessment horizon  $P = 3$ , FC sequence iteration  $S = 1600$  (*excluding* the initial Explore iterations), learning rate  $\alpha = 0.0005$ , and  $\kappa \in [3]$ .

### A.6.2 ACCURACY AND EFFICIENCY – PERFORMANCE RESULTS

Algorithm 1 is initiated after the experience buffer is filled for Explore search iterations. Table 3 illustrates the results obtained. Similar to 39-bus system, we observe that Algorithm 1 with a greater  $\kappa$  discovers FC sequences with larger accumulated TLLs and discovers more number of risky FCs leading to superior accuracy metrics for larger  $\kappa$ . For instance the average regret of Algorithm 1 with  $\kappa = 3$  is  $321.90 \times 10^3$  MWs and it is 0.4% lower the average regret when  $\kappa = 1$ . Similarly, the average precision when  $\kappa = 3$  is 11.7% higher that of  $\kappa = 1$ . Figure 4 illustrate the average accuracy metric for Algorithm 1 as a function of  $s \in [S]$ . A comprehensive discussion of additional performance results is provided in [42].

### A.6.3 COMPARISON WITH BASELINES

We perform comparisons with the two baselines approaches in [15], i.e., the PFW + RL and PFW + RL + TE. For the PFW + RL + TE approach, we first run their proposed  $Q$ -learning-based approach offline, for a loading condition of  $1.0 \times \text{base\_load}$  (bringing in the prior knowledge) for  $S = 5000$  iterations and store its extensive  $Q$ -table to run the PFW + RL + TE approach in real-time for the considered loading condition of  $0.6 \times \text{base\_load}$ , signifying a transition from  $1.0 \times \text{base\_load}$  to an extension to the current system loaded at  $0.6 \times \text{base\_load}$ . We set the parameters and hyper-parameters as described in Section A.6.1.

**Comparison under Unbounded Computational Budget** Table 3 compares the accuracy metrics for  $S = 1600$ , showing that Algorithm 1 consistently outperforms both the other baseline approaches. Figure 4 shows how the average regret scales as a function of search iterations  $s \in [S]$ .

**Comparison under Bounded Computational Budget** We next evaluate all the above approaches considering a run-time computational budget of five minutes, averaged over 25 MC iterations, and Table 4 illustrates the relative performance. All the observations corroborate those observed for the 39-bus system.

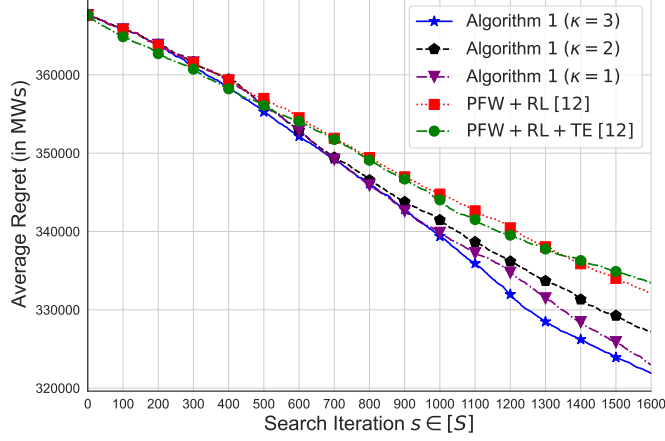


Figure 4:  $\text{Regret}(s)$  versus  $s$  for the IEEE-118 bus system.

Algorithm	Range for Accumulative TLL $\sum_{i=1}^S \text{TLL}(\mathcal{V}_i)$ (in GWs)	Range for $\text{Regret}(S)$ (in GWs)
Algorithm 1 ( $\kappa = 3$ )	$45.73 \pm 16\%$	$321.90 \pm 2.3\%$
Algorithm 1 ( $\kappa = 2$ )	$40.42 \pm 17\%$	$327.22 \pm 2.4\%$
Algorithm 1 ( $\kappa = 1$ )	$44.62 \pm 12\%$	$323.02 \pm 1.7\%$
PFW + RL + TE [15]	$34.23 \pm 2.7\%$	$333.40 \pm 0.28\%$
PFW + RL [15]	$35.50 \pm 2.4\%$	$332.14 \pm 0.25\%$

Table 3: Performance comparison for the IEEE-118 test system.

Algorithm	Average No. of FC Sequences $S$ Discovered	Range for Accumulative TLL $\sum_{i=1}^S \text{TLL}(\mathcal{V}_i)$ (in GWs)	Range for $\text{Regret}(S)$ (in GWs)
Algorithm 1 ( $\kappa = 3$ )	186	$11.46 \pm 16.2\%$	$175.63 \pm 3.4\%$
Algorithm 1 ( $\kappa = 2$ )	239	$18.32 \pm 18\%$	$212.22 \pm 4.1\%$
Algorithm 1 ( $\kappa = 1$ )	301	$19.86 \pm 13\%$	$256.75 \pm 2.6\%$
PFW + RL + TE [15]	527	$18.56 \pm 3\%$	$436.22 \pm 0.98\%$
PFW + RL [15]	522	$17.98 \pm 3\%$	$439.13 \pm 1.1\%$

Table 4: Performance comparison for a computational time of 5 minutes for the IEEE-118 bus test system.