Reinforcement Learning for Cyber-Physical Security Assessment of Power Systems

Xiaorui Liu, Charalambos Konstantinou
Department of Electrical and Computer Engineering, FAMU-FSU College of Engineering
Center for Advanced Power Systems, Florida State University
Tallahassee, FL, USA
E-mail: {xliu9, ckonstantinou}@fsu.edu

***Abstract*—The protection of power systems is of paramount significance for the supply of electricity. Contingency analysis allows to access the impact of power grid components failures. Typically, power systems are designed to handle $N-2$ contingencies. Existing algorithms mainly focus on performance and computational efficiency. There has been little effort in designing contingency methods from a cybersecurity perspective. To address this limitation, we study contingency analysis in the context of power system planning and operation towards cyber-physical security assessment. The proposed methodology considers attackers transitions in the network based on the $N-2$ critical contingency pairs. We develop an online reinforcement Q-learning scheme to solve a Markov decision process that models adversarial actions. In this model, the adversary aims to maximize the cumulative reward before making any action and learns adaptively how to optimize the attack strategies. We validate and test the algorithm on eleven literature-based and synthetic power grid test cases.**

***Index Terms*—Contingency analysis, cybersecurity, power systems security, reinforcement learning.**

I. Introduction

During the last years, power systems are transitioning towards a more monitored and managed grid. The newly formed smart grid infrastructure offers significant benefits in terms of reliable and efficient operation. This is driven by the integration of information and communication technologies into grid devices, services, and markets. While such integration is important towards a fully smart grid, it has also expanded the threat landscape, making the system more vulnerable to cyber-attacks [1]. The recent attack incidents of electric utilities in Ukraine demonstrate the implications that sophisticated cyber-attacks can cause on the critical infrastructure. In 2015, attackers compromised control systems and infected software with malicious code able to trip circuit breakers and cause a power outage. A year later, adversaries using grid sabotaging malware (CrashOverride) shut control relays and plunged the Kiev region into darkness for hours [2].

Due to the increased interdependency of cyber and physical components of modern power systems, cyber-threats continue to arise within the different layers of the infrastructure. Therefore, it is vital for both safety and economic reasons that the system is operating in the normal state when components fail. The ability of the power system to "survive" after facing a set of changes and disturbances (contingencies) while providing uninterrupted customer service and without entering an emergency or a restorative state is defined as power system security. The identification process of whether the system remains in a secure or emergency state is called *power system security assessment*. This assessment is a three-step phase which involves: *i)* security monitoring, *ii)* contingency analysis, and *iii)* security control.

Security monitoring provides the system operating conditions. Contingency analysis generates the necessary information to system operators about the security of the system, i.e., it can be viewed as a "what-if" scenario simulator that assesses, produces, and ranks the impact of unscheduled events on a power system [3]. A contingency can be the failure or loss of system components (e.g., generators, substations, lines, transformers, etc.). Such loss or failure could be either planned (e.g., scheduled maintenance) or unplanned – unforeseen event (e.g., fault, cyber-attack, etc.). This phase of system assessment includes contingency screening and ranking: short-listing critical contingencies from all credible ones and ranking them based on their severity index. In the scenario which the power system is identified to be insecure, appropriate control actions are taken in order to restore the system's secure state.

Traditionally, power systems are designed to sustain a single component outage ($N-1$ criterion). For instance, the North America Electric Reliability Corporation (NERC) enforces strict power security standards that require system operators to satisfy the $N-1$ constraint. Due to the growing complexity of modern power systems (e.g., significant generation uncertainty, malicious cyber-threats, distributed generations, etc.) regulatory agencies also require operators to ensure system stability in the event of two or more contingencies k: either k (near-)simultaneous losses [$N-k$ ($k \geq 2$) contingency] or consecutive losses [$N-1-1$ contingency] [4]. Despite the standardized $N-k$ criteria that enable power systems to move into a satisfactory state after k event(s), there always exists a number p of multiple contingencies, which power systems cannot withstand and will lead to cascading failures. Given the grid topology and power flows, the results of contingency analysis can be leveraged by adversaries towards identifying these p contingencies and target them in order to perform an attack and thus successfully compromise the system.

In this paper, we introduce a reinforcement learning process towards cyber-physical security assessment of power systems. We leverage existing fast pruning algorithms for contingency

screening and ranking to identify critical contingency pairs. Our focus is on $N-2$ contingencies due to their high occurrence possibility and their resulting severe impact on power systems [5]. For example, power supply reliability requirements for critical areas (e.g., leading lines of nuclear power plants) must satisfy the $N-2$ contingency criterion. We develop an online model-free reinforcement learning framework for the $N-2$ contingency problem in order to solve a discrete Markov decision process (MDP) that models the adversarial actions. Our automated iteration-based process demonstrates the optimal action which the attacker may take and associates the possible path with a corresponding reward.

The rest of the paper is structured as follows: Section II presents the related work on contingency analysis algorithms. The reinforcement learning methodology towards security assessment is described in Section III. Section IV demonstrates the simulation results on eleven benchmark systems of different scales. Finally, we conclude the paper and discuss future work in Section V.

II. Related work

In its simplest form, contingency analysis generates a power flow solution for each potential violation defined through the contingency list. The obtained information from the power flow study includes the voltage magnitude and phase angle at each bus and the active and reactive power flowing in each line. For each contingency event which is removed from the power system, the network model is simulated in order to calculate the corresponding power flow data for the remaining components. The results of each network solution are then used to determine if there exist any component limit violations (e.g., thermal ratings of transmission lines).

Regarding single contingency screening methods for real-time analysis, previous work has demonstrated a fast decoupled power flow algorithm [6]. Efficient contingency analysis schemes have also designed to detect violations of transmission lines power flow [7], [8]. For multiple contingencies, the main research effort is towards addressing the computational overhead of $N!/[k!(N-k)!]$ simulations [9]–[21]. Ranking and selection algorithms are the conventional methods that rank outages [9]–[11]. This can be performed, for example, using the performance index for line flows, voltage analysis, capacity, and power flow analysis. Other approaches are based on network topology analysis and nonlinear optimization heuristics [12], [13]. A concept of the line outage distribution factor (LODF) in [14], [16] is used as the basis for a fast $N-2$ algorithm that performs pruning of the contingency set [19], [20]. LODF is used to calculate the difference in the flow on one line caused by the outage of a second line. The complexity of the algorithm is $\mathcal{O}(N^2)$ comparable to the complexity of $N-1$ contingency problem.

Understanding the nature of threats in a smart grid environment helps to determine the appropriate security assessment and mitigation strategies [1]. Recently, it has been recognized that contingency analysis and assessment need to account for cyber-physical interactions. Ten *et al.* demonstrated the significance of cyber-based contingency analysis via an impact assessment of critical cyber-assets that collect historical load and topology information to identify critical substations [22]. In [23], a security-oriented risk management scheme generates cyber-physical security indices. A similar framework provides a unified formalism to model the cyber-physical system, rank the contingencies, and assess the potential impacts of events [24]. Xin *et al.* proposed a cyber-physical equivalent model for hierarchical control systems [25].

While there are a plethora of contingency analysis studies, there is a lack of literature that explores cyber-physical security assessment that addresses also the computational complexity of the problem. Hence, given an adversarial agent, it is essential to model adversary transitions to identify the optimal policy between the identified critical components. Reinforcement learning algorithms are known to be effective in optimally controlling uncertain environments. Such algorithms seek an action sequence in order to maximize some notion of cumulative reward via a trial-and-error approach. In this paper, we present an algorithm based on reinforcement learning which identifies the optimal action path of an adversary aiming to obtain a near-optimal discounted return that depends on the identified contingencies and their ranking.

III. Methodology

This section presents our proposed learning-based cyber-physical contingency analysis algorithm. The approach is based on the iterative bounding of LODFs and successive pruning of the contingency pair set [19], [20]. In this setting, the power system environment is modeled as a MDP in which the adversarial agent objective is to maximize his transition rewards between the pair of contingency sets. Specifically, the attacker could utilize a set of contingency pair candidates after the iterative pruning. The goal is to determine via active reinforcement learning an optimal transition policy in the set of the most critical contingency pair by choosing a series of actions. In the rest of the section, we provide the key relations necessary for the construction of our algorithm which is subsequently described.

A. Markov Decision Process (MDP)

A MDP provides a framework for modeling decision making under uncertainty. Depending on the model and the policy, the process seeks the optimal option for maximizing the expected sum of (discounted) rewards for a state. MDP can be modeled as a 4-tuple $(\mathscr{S}, \mathscr{A}, P_a, R_a)$ where $\mathscr{S}$ is a finite set of states and $\mathscr{A}$ is a finite set of attacker's action. P_a is the transition probability from the initial state s to state s' after action a is performed. R_a is the immediate reward received after P_a as the result of taking the decision a.

In the scenario which the probabilities or rewards are unknown, the problem is one of reinforcement learning. In this context, the Q-function of MDP corresponds the value of taking action a and then continuing optimally. $Q(s,a)$, shown in (1), is the expected utility starting at state s, taking action a, and hereafter, following an optimal strategy that provides

an aggregated reward. Thus, it is calculated via a cumulative function of the immediate reward $R(s, a)$. Discount factor $\gamma \in [0, 1]$ controls the importance of future rewards, i.e., the value of receiving reward R after n steps is $\gamma^n \cdot R$: this values immediate reward above postponed reward. For every state s, MDP corresponds to different a towards processing s: in our case, MDP evaluates the selection of the transitioning path for the most critical contingency pair.

$$Q(s, a) = \sum_{t=0}^{n} \gamma^t \cdot R(s, a) \tag{1}$$

B. The Q-Learning Algorithm

Q-learning is a semi-supervised reinforcement learning algorithm [26]. Its objective is to learn a policy and direct an agent towards an action a based on the model conditions. It is an off-policy, temporal difference reinforcement learning approach that approximates Q-value with Monte Carlo simulation. The core of the algorithm is shown in (2), where α is the learning rate ($0 \leq \alpha < 1$), $Q_{old}(s_t, a_t)$ is the cumulative value from the previous state, $R(s_t, a_t)$ is the immediate reward for the current state, and $\max_a Q(s_{t+1}, a)$ is the estimation of the optimal future Q-value. The algorithm is based on a value-based iteration using the old value and the new information. For each state-action pair (s, a), a Q-value is assigned in a Q-table. Initially, all the Q-values are set to zero.

$$\begin{aligned} Q_{new}(s_t, a_t) \leftarrow (1 - \alpha) \cdot Q_{old}(s_t, a_t) \\ + \alpha \cdot (R(s_t, a_t) + \gamma \cdot \max_a Q(s_{t+1}, a)) \end{aligned} \tag{2}$$

The optimal action a_t^* for a set of available actions A_t at state s_t in time t is shown in (3). a_t^* maximizes the total long-term reward Q instead of the instant reward R. In the scenario of multiple a_t^*, coordination mechanisms may break these ties.

$$a_t^* = \arg\max_{a \in A_t} Q(s_t, a_i) \tag{3}$$

Exploitation vs. Exploration Dilemma: The Q-function is utilized to determine an optimal path rather than enforcing actions to the agent. Based on the Q-value, an agent can: *i) explore* in order to construct a better estimate of the optimal Q-value, and *ii) exploit* the current knowledge at state s by taking action a to maximize $Q(s, a)$. There are two common approaches to trade-off exploration and exploitation. One way is follow "optimism in the face of uncertainty" by initializing Q-table to values that encourage exploration. Another method is the ϵ-greedy strategy which addresses the local optima problem during the learning, i.e., $a_t = a_t^*$ with probability $1 - \epsilon$ and select a random action a_t with probability ϵ.

C. Threat Model

Our threat model assumes an attacker would be able to leverage publicly available information using "open source intelligence" techniques to reconstruct a model of the target power system [3], [27], [28]. The attacker would be aware of the topological data of the system such as the location and status of circuit breakers. In addition, the attacker would be able to acquire the required data to perform power flow and contingency analysis: a structural connection between system buses, conductivities of lines, power injection data, and the acceptable range of power flow variation of the system buses and lines. Fig. 1 presents the flowchart of the attack process. Based on the collected data, the adversary performs a series of actions which would provide a reward value for each transition as well as an optimal future value for the overall set of actions. The process records the transition values in a table before the next state-iteration. The future Q-value, i.e., a cumulative reward function, is optimized by adjusting the actions that depend on the current reward and the estimated optimal future value. In this case, the updated Q-learning table records all the values for different states and actions that will be utilized for future decision making.

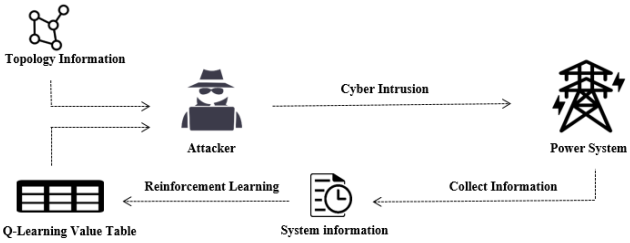

Figure 1. Overview of the reinforcement learning -based attack process.

D. Proposed Algorithm

Through Q-learning, we solve the MDP model of the power system network in which adversarial actions are considered. Our algorithm first leverages a LODF-based contingency pruning method to list the critical contingencies – overload lines [19]. The line overload condition can be written in the form of $A_{xy} \cdot B_{xc} + A_{yx} \cdot B_{yc} > 1$, where *x* and *y* are the outages lines considering the change of the flow on some arbitrary line *z* and *c* is the possible constraint. Matrix A_{xy} can be calculated by $A_{xy} = (1 + L_{xy} \cdot f_y / f_x)/(1 - L_{yx} \cdot L_{xy})$, where L is the LODF. For example, $L_{xy} = (f'_x - f_x)/f_y$ relates the change of the flow through line x from f_x to f'_x with the flow f_y through line y before the outage (line trip). Matrix B_{xc} is given by $B_{xc} = f_x \cdot L_{zx}/(f_z^{critical} \pm f_z)$ where $f_z^{critical}$ is the bound value and the $+, -$ signs correspond to the conditions $f'_z < -f_z^{critical}$ and $f'_z > f_z^{critical}$, respectively.

After sorting the overload lines, Algorithm 1 shows the pseudo-code to estimate the movement of the adversary between the identified critical contingency pairs. C is a set which includes all the $N - 2$ overload contingency pairs and specifically, C_p contains the most critical contingency pair (*lines 1-2*). If C_p is not an empty set, there must be a one-to-one correspondence between $C_p(s)$ and $C_p(s')$, where s is the initial attacker state in the network model and s' is the target state based on the contingency pairs list. ΔC_p is determined

Algorithm 1 Proposed algorithm

1: $C = \{A_{xy} \cdot B_{xc} + A_{yx} \cdot B_{yc} > 1\}$
2: $C_p = Max\{C\}$
3: If $C_p \neq \varnothing$
4: $\quad \Delta C_p \leftarrow C_p(s') - C_p(s)$ ▷ Physical Impact
5: **end**
6: $P_1 = min(\Delta C_p)$
7: $P_2 = W(s) \cdot (\Delta C_p)$
8: $P = \sum_{K=1}^{m} n_k P_k, \sum_{k=1}^{m} n_k = 1$
9: **Repeat**
10: For $a \in \mathscr{A}$ do
11: $S_i = \max_{a \in \mathscr{A}} \gamma \cdot \sum_{s' \in \mathscr{S}} P \cdot [\Delta C_p + S_i']$ ▷ Security Index
12: $R = \sum_{s' \in \mathscr{S}} P \cdot [\Delta C_p + S_i'])$ ▷ Reward
13: Find all $a' = N(a)$
14: Calculate $\max_a Q(s', a')$
15: $Q(s,a) = (1-\alpha) \cdot Q(s,a) + \alpha \cdot (R + \gamma \cdot \max_a Q(s', a'))$ ▷ Q
16: **until** $s = s'$
17: **return** Q

based on the initial and final-target states in order to formulate the physical impact on the system (*lines 3-5*).

Based on the calculated physical impact ΔC_p, the adversary movement needs to be defined. The cyber-attack path defines P as the transition probability. The adversary can use the shortest path P_1 which spans over a minimum number of nodes or consider an "easier-for-transition" path P_2 that may progress over a higher number of system nodes. P_1 is obtained by finding the minimum number nodes between $C_p(s)$ and $C_p(s')$ that constitute a transition path. The calculation of P_2 depends on a weight function $W(s)$, which is a connectivity matrix formulated based on the reactance of the system transmission lines. The probability of the attacker path selection is composed by $P = \sum_{K=1}^{m} n_k P_k$, where $\sum_{k=1}^{m} n_k = 1$. In this work, we consider $m = 2$ while n_k is a random variable following the continuous uniform distribution over the set $[0, 1]$ (*lines 6-8*).

After estimating the physical impact and the transition probability, the next step is to find the access difficulty of each different path. Depending on the adversary actions, each initial state can change. S_i calculates the overall security index for physical impact C_P. $\mathscr{A}$ includes the set of all actions $a \in \mathscr{A}$ for which the value of S_i can be known. It is a dynamic function which includes the security index for state s' (*line 11*). Furthermore, R is used to represent the immediate reward of attack and depends on the different physical impact (ΔC_p) and security index (S_i) selected by the adversary (*line 12*).

In addition to using the immediate reward R, another important parameter for calculating the Q-learning value of MDP is the estimate of the optimal future value. In the current state, the first step in achieving the optimal future value is to find all possible actions a' for next state s' after action a occurs. This is shown as $a' = N(a)$ (*line 13*). The selected

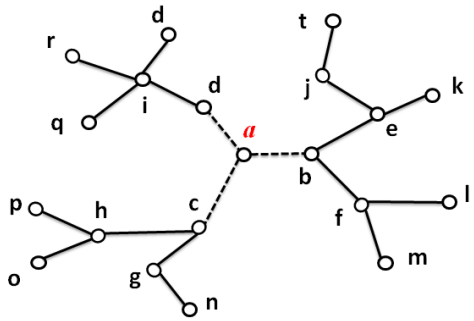

Figure 2. Network graph for a toy example.

action a' is based on the selection of the highest reward. The maximum Q-value for performing action a' in the next step (s') will be taken into consideration for the optimal future value (*line 14*). In our exploitation, we follow an ϵ-greedy strategy in which a_t is always selected as a_t^* in order to maximize reaching the target state ($\epsilon = 0$). Fig. 2 shows the network graph for a toy example in order to illustrate the process. For example, if a is the current state, a' can be calculated as $a' = N(a) = (b, c, d)$. If choosing action b could gain a better reward rather than c and d, then the value of selecting action b will be used for the optimal Q-future value. If b is the current state, then $b' = N(b) = (e, f)$. This process is repeated for every different state.

Based on the immediate reward and optimal future values, the Q-learning process is initiated and creates the Q-table (*line 15*). Every random action will keep running repeatable for updating the Q-table. Through iterations (*lines 16-17*), the Q-table will display all the cumulative reward of various paths and present the one that would reach optimally and easiest the target state. The set of actions that need to be taken at each step will form the attack path.

IV. Case Studies

In order to validate and evaluate the proposed algorithm, the simulation tests for both literature-based and synthetic power grid test cases were performed on a 64-bit machine with an Intel Core i7-8700 3.2 GHz CPU, 16.00 GB of memory, using the MATPOWER package of MATLAB.

The algorithm first identifies all the $N-2$ contingency pairs using the described LODF-based pruning method. Then, the reinforcement learning scheme is applied to generate the possible attacker paths in the network model of the power system test case. In order to understand the effectiveness of our approach, we compare the results of the transition path of our learning algorithm with the shortest and weighted path. We also evaluate the performance in terms of timing overhead of the proposed algorithm compared with simply performing the LODF-based contingency pruning method to list the critical overload lines of the system [19].

Our results include test cases of various node sizes. Table I shows the transition route for the pair of nodes $\{2196, 2320\}$ and $\{34, 222\}$, i.e., for the transmission line pair $2196-2320$ and $34-222$, which is ranked as the most severe one from

all 9964 contingency pairs for the Polish 3012 bus system (winter 2007-08 evening peak). The results are shown for each transition probability method between the contingency line pair. Out of all transition paths, the route between bus 2320 to bus 34 is identified to be the Q-optimal (with seven steps) based on our proposed algorithm.

Table II shows the results of the proposed learning approach including the number of contingency pairs for each test case, the number and order of nodes for the most critical pair in the transition path, as well as the performance overhead. For each case study, we present the transition path based on Algorithm 1, between the buses of the system $\{a, b\}$ and $\{m, n\}$ when ab and mn is the most critical contingency pair of edges (lines) in the system. For example for the ACTIVSg2000: 2000-bus synthetic grid on footprint of Texas, our approach lists four path routes between the nodes of the system that form the most critical contingency line pair $\{1791, 1593\}$ and $\{1735, 1660\}$. Out of all the possible transitions paths, the transition route between bus 1593 and 1660 is the most optimal with only four transitions.

There are eleven in total simulation results from different test cases. Despite the fact that the shortest and the weighted path provide the attacker with a transition route, these paths do not estimate the optimal movement of this transition. For test cases with a large number of buses, the overhead of the proposed algorithm does not exceed 1% compared with the time needed to identify the critical contingency pairs. For instance, the contingency analysis algorithm used for this paper requires 2365.36 $secs$ to list and rank all contingency pairs for the Polish 3012 bus system (winter 2007-08 evening peak). The time required to run our proposed reinforcement learning algorithm for this case is 1.125 $secs$, that is 0.048% in terms of performance overhead. On average for all tested bus system cases, the running time of our algorithm is 1.064 $secs$.

Table I
POLISH 3012 BUS SYSTEM - WINTER 2007-08 EVENING PEAK.

A/A	# Trans. Lines	Method	Path (system buses)
1	10	$\mathcal{W}$	2196, 2320, 170, 169, 155, 154, 27, 37, 40, 34
	8	$\mathcal{S}$	2196, 2320, 170, 154, 27, 37, 40, 34
	8	$\mathcal{Q}$	2196, 2320, 170, 154, 27, 37, 40, 34
2	10	$\mathcal{W}$	2196, 2320, 170, 169, 192, 225, 226, 221, 220, 222
	8	$\mathcal{S}$	2196, 2320, 170, 191, 226, 221, 220, 222
	9	$\mathcal{Q}$	2196, 2320, 170, 154, 27, 37, 40, 34, 222
3	9	$\mathcal{W}$	2320, 170, 169, 155, 154, 27, 37, 40, 34
	7	$\mathcal{S}$	2320, 170, 154, 27, 37, 40, 34
	7	$\mathcal{Q}$	2320, 170, 154, 27, 37, 40, 34
4	9	$\mathcal{W}$	2320, 170, 169, 192, 225, 226, 221, 220, 222
	7	$\mathcal{S}$	2320, 170, 191, 226, 221, 220, 222
	8	$\mathcal{Q}$	2320, 170, 154, 27, 37, 40, 34, 222
where $\mathcal{W}$, $\mathcal{S}$, $\mathcal{Q}$ refer to the weighted, shortest, and Q-learning-based Algorithm 1, respectively.			

V. CONCLUSIONS AND FUTURE WORK

In this paper, we present a security assessment algorithm for power systems in which an adversarial agent identifies the optimal transition policy through the system nodes. The reinforcement learning approach generates the optimal transition path for an attacker following the ranking results from contingency analysis. This is achieved by solving a MDP that models adversarial actions in an automated iteration-based method which is based on Q-learning. Simulation results with eleven test cases demonstrate the effectiveness of the algorithm. Future work on the topic will include modeling the system using a partially observable MDP. Also, we will develop and test our scheme using a digital real-time simulator in order to capture the transient and dynamic behavior of complex electrical power systems in real-time.

ACKNOWLEDGEMENTS

This work was partially supported by the FSU FYAP grant program.

REFERENCES

[1] C. Konstantinou and M. Maniatakos, "Security analysis of smart grid," *Chapter 15 – Communication, Control and Security Challenges for the Smart Grid*, vol. 2, p. 451, 2017.

[2] BBC News, "Ukraine power cut 'was cyber-attack'," [Online]. Available: https://www.bbc.com/news/technology-38573074, 2017.

[3] A. Keliris *et al.*, "Open source intelligence for energy sector cyberattacks," in *Critical Infrastructure Security and Resilience*. Springer, 2019, pp. 261–281.

[4] NERC, "Standard TPL-001-1," [Online]. Available: https://www.nerc.com/files/TPL-003-0.pdf.

[5] S.-E. Chien *et al.*, "Automation of contingency analysis for special protection systems in taiwan power system," in *2007 International Conference on Intelligent Systems Applications to Power Systems*. IEEE, 2007, pp. 1–6.

[6] G. Ejebe, R. Paliza, and W. Tinney, "An adaptive localization method for real-time security analysis," *IEEE transactions on power systems*, vol. 7, no. 2, pp. 777–783, 1992.

[7] V. Brandwajn, "Efficient bounding method for linear contingency analysis," *IEEE Transactions on Power Systems*, vol. 3, no. 1, pp. 38–43, 1988.

[8] U.S. FERC, "Mandatory reliability standards for the bulk-power system," [Online]. Available: https://www.ferc.gov/whats-new/comm-meet/2007/031507/E-13.pdf.

[9] B. Stott, O. Alsac, and F. Alvarado, "Analytical and computational improvements in performance-index ranking algorithms for networks," *International Journal of Electrical Power & Energy Systems*, vol. 7, no. 3, pp. 154–160, 1985.

[10] G. Ejebe, H. Van Meeteren, and B. Wollenberg, "Fast contingency screening and evaluation for voltage security analysis," *IEEE Transactions on Power Systems*, vol. 3, no. 4, pp. 1582–1590, 1988.

[11] M. K. Enns, J. J. Quada, and B. Sackett, "Fast linear contingency analysis," *IEEE Transactions on Power Apparatus and Systems*, no. 4, pp. 783–791, 1982.

[12] Q. Chen and J. D. McCalley, "Identifying high risk n-k contingencies for online security assessment," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 823–834, 2005.

[13] T. Guler and G. Gross, "Detection of island formation and identification of causal factors under multiple line outages," *IEEE Transactions on Power Systems*, vol. 22, no. 2, pp. 505–513, 2007.

[14] C. M. Davis and T. J. Overbye, "Multiple element contingency screening," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1294–1301, 2011.

[15] M. J. Eppstein and P. D. Hines, "A "random chemistry" algorithm for identifying collections of multiple contingencies that initiate cascading failure," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1698–1705, 2012.

Table II
RESULTS.

A/A	Test Case	# Cont. Pairs	# Trans. Lines	Path (system buses)	Overhead
1	IEEE 39-bus system	71	5	16, 15, 14, 13, 10	59.343%
			4	16, 15, 14, 13	
			6	21, 16, 15, 14, 13, 10	
			5	21, 16, 15, 14, 13	
2	UIUC 150-bus system	442	9	140, 27, 106, 102, 96, 101, 144, 95, 137	27.6327%
			8	140, 27, 106, 102, 96, 101, 144, 95	
			8	39, 9, 23, 96, 101, 144, 95, 137	
			7	39, 9, 23, 96, 101, 144, 95	
3	SouthCarolina bus system: ACTIVSg500 (500 buses)	330	9	322, 58, 57, 440, 297, 65, 80, 407, 408	23.5965%
			10	322, 58, 57, 440, 297, 65, 80, 407, 408, 212	
			8	58, 57, 440, 297, 65, 80, 407, 408	
			9	58, 57, 440, 297, 65, 80, 407, 408, 212	
4	Synthetic grid on footprint of Texas: ACTIVSg2000 (2000 buses)	977544	10	1791, 1593, 1592, 1446, 1817, 1750, 1749, 1712, 1713, 1735	0.0256%
			5	1791, 1593, 1592, 1755, 1660	
			9	1593, 1592, 1446, 1817, 1750, 1749, 1712, 1713, 1735	
			4	1593, 1592, 1755, 1660	
5	Polish system winter 1999-2000 peak (2383 buses)	15881	7	381, 344, 346, 10, 5, 6, 310	0.0842%
			6	381, 344, 346, 10, 5, 6	
			8	314, 381, 344, 346, 10, 5, 6, 310	
			7	314, 381, 344, 346, 10, 5, 6	
6	Polish system summer 2004 peak (2736 buses)	1680	15	2063, 1966, 2435, 1929, 133, 146, 145, 164, 139, 168, 169, 172, 180, 152, 2153	0.2045%
			14	2063, 1966, 2435, 1929, 133, 146, 145, 164, 139, 168, 169, 172, 180, 152	
			14	1966, 2435, 1929, 133, 146, 145, 164, 139, 168, 169, 172, 180, 152, 2153	
			13	1966, 2435, 1929, 133, 146, 145, 164, 139, 168, 169, 172, 180, 152	
7	Polish system summer 2004 off-peak (2737 buses)	463	14	2562, 2092, 147, 134, 20, 19, 26, 28, 123, 121, 120, 63, 64, 47	0.9488%
			13	2562, 2092, 147, 134, 20, 19, 26, 28, 123, 121, 120, 63, 64	
			16	2702, 2699, 2698, 204, 202, 201, 190, 192, 25, 28, 123, 121, 120, 63, 64, 47	
			15	2702, 2699, 2698, 204, 202, 201, 190, 192, 25, 28, 123, 121, 120, 63, 64	
8	Polish system winter 2003-04 evening peak (2746 buses)	4344	13	483, 403, 24, 33, 27, 29, 28, 123, 121, 120, 64, 65, 48	0.0863%
			12	483, 403, 24, 33, 27, 29, 28, 123, 121, 120, 64, 65	
			12	403, 24, 33, 27, 29, 28, 123, 121, 120, 64, 65, 48	
			11	403, 24, 33, 27, 29, 28, 123, 121, 120, 64, 65	
9	Polish system winter 2003-04 off-peak (2746 buses)	918	12	2109, 2086, 1981, 138, 139, 164, 163, 82, 127, 93, 91, 131	0.3199%
			13	2109, 2086, 1981, 138, 139, 164, 163, 82, 127, 93, 91, 131, 130	
			11	2086, 1981, 138, 139, 164, 163, 82, 127, 93, 91, 131	
			12	2086, 1981, 138, 139, 164, 163, 82, 127, 93, 91, 131, 130	
10	Polish system winter 2007-08 evening peak (3012 buses)	9964	8	2196, 2320, 170, 154, 27, 37, 40, 34	0.0256%
			9	2196, 2320, 170, 154, 27, 37, 40, 34, 222	
			7	2320, 170, 154, 27, 37, 40, 34	
			8	2320, 170, 154, 27, 37, 40, 34, 222	
11	Polish system summer 2008 morning peak (3120 buses)	44247	15	934, 928, 62, 61, 137, 139, 138, 130, 145, 142, 119, 120, 110, 109, 1530	0.0372%
			14	934, 928, 62, 61, 137, 139, 138, 130, 145, 142, 119, 120, 110, 109	
			16	930, 934, 928, 62, 61, 137, 139, 138, 130, 145, 142, 119, 120, 110, 109, 1530	
			15	930, 934, 928, 62, 61, 137, 139, 138, 130, 145, 142, 119, 120, 110, 109	

[16] C. Davis and T. Overbye, "Linear analysis of multiple outage interaction," in *System Sciences, 42nd Hawaii International Conference on*. IEEE, 2009, pp. 1–8.

[17] V. Donde *et al.*, "Severe multiple contingency screening in electric power systems," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 406–417, 2008.

[18] D. Bienstock and A. Verma, "The n-k problem in power grids: New models, formulations, and numerical experiments," *SIAM Journal on Optimization*, vol. 20, no. 5, pp. 2352–2380, 2010.

[19] K. S. Turitsyn and P. Kaplunovich, "Fast algorithm for n-2 contingency problem," in *System Sciences (HICSS), 2013 46th Hawaii International Conference on*. IEEE, 2013, pp. 2161–2166.

[20] P. Kaplunovich and K. Turitsyn, "Fast and reliable screening of n-2 contingencies," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4243–4252, 2016.

[21] C. M. Rocco *et al.*, "Assessing the vulnerability of a power system through a multiple objective contingency screening approach," *IEEE Transactions on Reliability*, vol. 60, no. 2, pp. 394–403, 2011.

[22] C.-W. Ten, A. Ginter, and R. Bulbul, "Cyber-based contingency analysis," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3040–3050, 2016.

[23] C. Vellaithurai *et al.*, "Cpindex: cyber-physical vulnerability assessment for power-grid infrastructures," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 566–575, 2015.

[24] S. Zonouz *et al.*, "Socca: A security-oriented cyber-physical contingency analysis in power infrastructures," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 3–13, 2014.

[25] S. Xin *et al.*, "Cyber-physical modeling and cyber-contingency assessment of hierarchical control systems," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2375–2385, 2015.

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

[27] U.S. DHS and FBI, "Russian government cyber activity targeting energy and other critical infrastructure sectors," [Online]. Available: https://www.us-cert.gov/ncas/alerts/TA18-074A.

[28] C. Konstantinou, M. Sazos, and M. Maniatakos, "Attacking the smart grid using public information," in *2016 17th Latin-American Test Symposium (LATS)*. IEEE, 2016, pp. 105–110.