Resilient Transmission Hardening Planning in a High Renewable Penetration Era

Ali Bagheri, *Student Member, IEEE*, Chaoyue Zhao, *Member, IEEE*, Feng Qiu, *Senior Member, IEEE*, and Jianhui Wang, *Senior Member, IEEE*

***Abstract*—Hardening components in transmission systems is a practice to improve system resilience against possible disturbances caused by natural disasters. In a power system with a very high penetration of renewable energy, the system hardening will be further complicated by the uncertainty and variability of renewable energy. In this paper, we study the transmission line hardening planing problem in the context of probabilistic power flows injected by the high penetration of renewable energy. We assume that the probabilistic information of renewable energy is incomplete and ambiguous and propose a data-driven approach to approximate the renewable uncertainty sets. We then extend the $N-1$ security criteria to multiple simultaneous contingencies and seek to prepare a hardening plan for the worst-case scenarios. A two-stage data-driven stochastic model is formulated by considering the joint worst-case wind output distribution and transmission line contingencies. Then, we apply the Column-and-Constraints generation method to solve the proposed model. To test the effectiveness of the proposed approach, we conduct experiments on 24-bus and 118-bus test systems. We numerically show that the data-driven approach can effectively address the uncertainty ambiguity and the proposed approach can produce effective hardening plans that improve the system resilience.**

***Index Terms*—Transmission system resilience, data-driven stochastic programming, wasserstein metric, column-and-constraint generation.**

NOMENCLATURE

A. Sets

$\mathcal{B}$	Index set of all buses.
$\mathcal{B}_i$	Index set of all buses directly connected to bus i.
$\mathcal{E}$	Index set of transmission lines.
$\mathcal{E}_A$	Index set of attacked transmission lines.
$\mathcal{E}_H$	Index set of hardened transmission lines.
$\mathcal{T}$	Index set of load blocks.
N	Index set of wind output scenarios.

B. Parameters

L_{it}	Load shedding cost at bus i for load block t.
H_{ij}	Investment cost to harden transmission line (i,j).
U	The maximum number of lines affected by natural disasters simultaneously.
F_{ij}	Flow capacity of transmission line (i,j).
C_i	Generation unit capacity at bus i.
X_{ij}	Reactance of transmission line (i,j).
$\theta_i^{\min}$	Phase angle lower limit at bus i.
$\theta_i^{\max}$	Phase angle upper limit at bus i.
d_{it}	Demand at bus i for load block t.
w_{it}	Renewable energy output at bus i, load block t.

C. Decision Variables

z_{ij}	Binary decision variable to indicate whether transmission line (i,j) is hardened ($z_{ij}=1$) or not ($z_{ij}=0$).
u_{ij}	Binary decision variable to indicate whether transmission line (i,j) is attacked ($u_{ij}=0$) by natural disaster or not ($u_{ij}=1$).
p_{it}	Power generation at bus i for load block t.
$f_{ij,t}$	Power flow from bus i to bus j on transmission line (i,j) for load block t.
θ_{it}	Phase angle at bus i for load block t.
s_{it}	Load shedding at bus i for load block t.

Manuscript received August 9, 2017; revised December 3, 2017, March 16, 2018, and July 3, 2017; accepted July 18, 2018. Date of publication October 1, 2018; date of current version February 18, 2019. The work of A. Bagheri and C. Zhao are supported by the National Science Foundation (NSF-1662589 and NSF-1610935). The work of F. Qiu is supported by the Advanced Grid Modeling Program at the U.S. Department of Energy Office of Electricity under the contract number DE-AC02-06CH11357. The work of J. Wang is supported by the U.S. Department of Energy Office of Electricity. Paper no. TPWRS-01236-2017. *(Corresponding author: Feng Qiu.)*

A. Bagheri and C. Zhao are with the School of Industrial Engineering and Management, Oklahoma State University, Stillwater, OK 74074 USA (e-mail: ali.bagheri@okstate.edu; chaoyue.zhao@okstate.edu).

F. Qiu is with the Energy Systems Division, Argonne National Laboratory, Lemont, IL 60439 USA (e-mail: fqiu@anl.gov).

J. Wang is with the Department of Electrical Engineering, Southern Methodist University, Dallas, TX 75205 USA (e-mail: jianhui@smu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2018.2872893

I. INTRODUCTION

NATURAL disasters, such as hurricane, wildfire, flooding, etc., have become more frequent in the recent years [1]. Natural disasters often disrupt power system operations and cause service interruptions, ranging from short-term service losses to large-area extended outages. For example, according to a report by the President's Council of Economic Advisers and the U.S. Department of Energy [2], power outages caused by severe weather conditions constitute 58% of all outages and 87% of outages affecting 50000 or more electricity customers, during the period from 2003 to 2012. It is estimated that weather incurred outages have cost the United States 60 billion USD annually [2]. These facts indicate the urgency of improving power system resilience against extreme weather events. To mitigate the natural disaster related risks and improve power grid resilience, many research works along with optimization tools focus on three main thrusts in the process of the power grid

in reacting to the nature disasters: 1) pre-disaster system hardening and investment (e.g., [3]–[13]), 2) emergency responses and corrective actions during or right after disasters (e.g., [14]–[17]), and 3) self-healing, rapid system restoration and damage assessment after disasters (e.g., [18]–[21]). Hardening, as one of the most effective activities to increase the power system resilience, is defined as any physical change (such as undergrounding power lines, vegetation management, pole reinforcing, etc.) to the power system infrastructure to make it less vulnerable to damage from severe weather conditions [22], [23].

One of the most commonly used security criteria for daily operations is the $N-1$ contingency (e.g., [24], [25]), where the system can continue operations without load shedding under any single component failure. A more stringent but less used criterion is the $N-k$ contingency, in which the system is required to sustain simultaneous failures of k electrically connected components [26]. While these security criteria effectively represent system operators' concerns for daily operations, they do not capture the possible contingencies in extreme weather conditions or natural disaster occurances, in which multiple components that are not electrically connected could fail simultaneously [27]. In this work, we generalize the $N-k$ contingencies to any k simultaneous component failures and aim to prepare for the worst k failures. Finding a defensive strategy against the worst case among a set of adversary scenarios is often modeled as a bi-level interdiction framework, which is known as the attacker-defender (AD) model (e.g., [3], [5], [6]). In this context, the attackers are extreme weather events (natural disasters) trying to cause the most severe damage to system operations; the defenders are the re-dispatch actions that minimize the damages by redistributing power flows. However, the AD model helps to find near-optimal but not the optimal protection plan against disruptive events, because it only seeks for the most critical set of assets, and hardening these critical assets is not necessarily the optimal protection plan [7]. To obtain the optimal hardening decisions, the tri-level attacker-defender-attacker (DAD) model was initially proposed by Brown et al. [8]. The DAD model, which is an extension of the AD model, includes two interacting agents (attacker and defender) in three levels (see, e.g., [9]–[12]). In the first level, the defender makes hardening decisions with a limited protection budget before disruptions. In the second level, the attacker disrupts the system to make the defender suffer from the highest cost or largest load shedding. In the third level, the defender aims to minimize the system cost against disruptions by taking corrective actions through re-dispaching the power output.

Besides the vulnerability of power system to nature disasters, the increasing renewable energy integration to the power system brings another significant challenge for independent system operators (ISOs) to build a sustainable power system infrastructure. As renewable penetration continues to grow (e.g., it is predicted that 20% of nation's electricity is generated by wind energy by 2030 [28]), to mitigate the risks of disruptions caused by the intermittent nature of renewable energy, the ISOs are responsible for preparing the grid to enhance the system resilience with the uncertain renewable energy output. To capture the renewable energy uncertainty into optimization models, stochastic programming (SO) (see, e.g., [29]–[32]) and robust optimization (RO) (see, e.g., [33]–[35]) approaches have been studied extensively. However, SO and RO approaches have shown disadvantages in practice. SO approaches can be unreliable due to the blind assumption of probability distribution of the random parameter. In addition, SO becomes computationally challenging for a large number of random parameter scenarios. Also, RO approaches can be too conservative due to the consideration of the worst-case scenario of the uncertain parameter, which happens rarely.

Another challenge for these methodologies to be practical is that, an accurate/complete knowledge on the probability distribution can hardly be obtained. Recently, distributionally robust and data-driven optimization (DRO) (see, e.g., [36]–[38]) has been applied to power system operations. The advantages of DRO is that it can handle uncertainties with partial information about the probability distribution. DRO approaches have been successfully developed and implemented to solve power system optimization problems under uncertainty (see, e.g., [39]–[42]). In this approach, by learning from a set of historical data or moment information for the random parameter, an ambiguity set for the unknown probability distribution of the random parameter is constructed. Then, the objective is to minimize the total cost under the worst-case distribution scenario within the constructed ambiguity set. Due to the considerations of the worst-case distribution, DRO leads to risk-averse and conservative solutions, as compared to traditional stochastic approaches. However, the DRO in general is less conservative than the traditional robust approaches [37]. That is because, DRO takes advantage of data information to build an ambiguity set of distribution and considers the worst-case distribution in the ambiguity set to keep robustness, while RO ignores the probability of random parameter scenarios, which usually leads to an unnecessarily high average cost. We will also numerically show this fact in our case study.

In this study, we develop a defender-attacker-defender-based transmission system hardening planning (TSHP) model under both random disruptions (natural disasters) and uncertain wind power generation. Due to the availability of a considerable amount of historical data for wind power output to the Independent System Operators (ISOs), we deploy the DRO approach to formulate the wind output uncertainty. That is, we allow the ambiguity of the probability distribution for the wind output, and construct the ambiguity set for the unknown distribution. More specifically, to build the ambiguity set, we use the Wasserstein metric as a probability measure [43], [44], which also has many applications in transportation theory [45]. The built transmission system hardening planning model is a two-stage model that deals with the hardening decisions in the first stage and the re-dispatching decision in consideration with the worst-case disruption scenario and the worst-case wind output probability distribution in the second stage. This two-stage model can be reformulated as a tractable formulation that can be solved efficiently by using decomposition approaches. The contributions of this research are summarized as follows: A planning tool for power systems with high renewable generation capacity is developed. This tool, which considers both renewable generation uncertainty and multiple random disruptions, helps power system operators to efficiently allocate protective resources in order to reduce the vulnerability of the power transmission

system against multiple transmission line contingencies caused by natural disasters or terrorist attacks, as well as maintain power system reliability with a large penetration of renewable energy. In addition, the proposed framework can efficiently utilize data information to reduce the conservativeness, and moreover, it can be solved efficiently with a tractable reformulation.

The remaining parts of this paper are organized as follows: In Section II, we describe how to use the historical data and construct the confidence set. In Section III, we develop a two-stage data-driven defender-attacker-defender model under both natural disaster and wind power uncertainties. In Section IV, the proposed decomposition framework and the solution algorithm are presented. In Section V, numerical results are discussed. Finally, the research is concluded in Section VI.

II. Data-Driven Modeling and Confidence Set Construction

In this research, we allow the probability distribution of wind output to be ambiguous because a particular distribution assumption for wind power output can be biased from the actual one. We construct an ambiguity set for the true probability distribution for wind output, which is centered at the reference distribution that is learned from a given historical data set.

First, we discuss the way to use the historical data and to obtain the reference distribution to estimate the true distribution. Given a set of historical data points, we let the histogram of data be the reference distribution of wind power output. Without loss of generality, we assume the random wind power output $w(\xi)$ to be bounded above and below, within a supporting space $\mathcal{W}$. We partition $\mathcal{W}$ into N bins $B_1, \ldots, B_N$, so that $\mathcal{W} = \bigcup_{n=1}^{N} B_n$. Given a set of historical data with size S, we obtain the reference distribution $\hat{\mathrm{P}} = (\hat{\mathrm{p}}^1, \hat{\mathrm{p}}^2, \ldots, \hat{\mathrm{p}}^N)$, where $\hat{\mathrm{p}}^n = S_n/S, \forall n = 1, 2, \ldots, N$, and S_n denotes the frequency of data samples in bin n. Notice here the size of bins, i.e., N affects both time complexity and solution accuracy. Actually in our data-driven modeling approach, N is the only parameter that affects the computational complexity. That is, as N increases, the computational complexity increases. However, the increase of N also leads to the increase of solution accuracy. Therefore, there is a trade-off between increasing accuracy and improving computational efficiency on the choice of N.

Second, we use the reference distribution to construct a confidence set (or ambiguity set) for the true distribution of wind output with confidence level β. Note that the reference distribution $\hat{\mathrm{P}}$ is inherently different from the true distribution P. To measure the difference between P and $\hat{\mathrm{P}}$, i.e., $\mathfrak{D}(\mathrm{P}, \hat{\mathrm{P}})$, we use the Wasserstein metric as the probability measure [43], [44]. Accordingly, $\mathfrak{D}(\mathrm{P}, \hat{\mathrm{P}})$ is defined as shown below:

$$\mathfrak{D}_{\mathrm{w}}(\mathrm{P}, \hat{\mathrm{P}}) := \inf_{\mathbb{Q}} \left\{ E_{\mathbb{Q}}[\mathrm{d}(w, \hat{w})] : \mathrm{P} = \varphi(w), \hat{\mathrm{P}} = \varphi(\hat{w}) \right\}, \tag{1}$$

where w and $\hat{w}$ are random wind power output associated with the true distribution and the reference distribution, respectively. $\mathrm{d}(w, \hat{w})$ is the distance between random variables w and $\hat{w}$. $\mathbb{Q}$ denotes all joint distributions of w and $\hat{w}$ with marginal distributions P and $\hat{\mathrm{P}}$. $\varphi(.)$ denotes that P and $\hat{\mathrm{P}}$ are distribution functions.

Then, we can construct a distribution-based ambiguity set using the Wasserstein metric, as shown below:

$$\begin{aligned} \mathbb{D} &= \left\{ \mathrm{P} \in \mathfrak{P}_+ : \mathfrak{D}_{\mathrm{w}}(\mathrm{P}, \hat{\mathrm{P}}) \le \alpha \right\} \\ &= \left\{ \mathrm{P} \in \mathfrak{P}_+ : \inf_{\mathbb{Q}} \left\{ E_{\mathbb{Q}}[\mathrm{d}(w, \hat{w})] : \right. \right. \\ &\qquad \left. \left. \mathrm{P} = \varphi(w), \hat{\mathrm{P}} = \varphi(\hat{w}) \right\} \le \alpha \right\}, \end{aligned} \tag{2}$$

where $\mathfrak{P}_+$ represents the set of all probability distributions and α denotes the tolerance level of the distance, which depends on the confidence level β and the size of the historical data S. Under the Wasserstein metric, the relationship between α and S can be described by the following proposition (please see the Appendix for the proof):

Proposition 1: Given a set of historical data of size S, N bins and a supporting space $\mathcal{W}$ with diameter D, the convergence rate between P and $\hat{\mathrm{P}}$ under the Wasserstein metric is as follows:

$$Pr(\mathfrak{D}_{\mathrm{w}}(\mathrm{P}, \hat{\mathrm{P}}) \le \alpha) \ge 1 - 2N \exp(-4\alpha S/ND). \tag{3}$$

Accordingly, if the confidence level, i.e., the right-hand side of inequality (3), is set to be β, then we have

$$\alpha = \frac{ND}{4S} \log\left(\frac{2N}{1-\beta}\right). \tag{4}$$

Based on (4), as the size of historical data S increases, the value of α decreases, i.e., the distance between the reference distribution and the true distribution $\mathfrak{D}(\mathrm{P}, \hat{\mathrm{P}})$ becomes smaller, and $\hat{\mathrm{P}}$ converges to P.

We denote the central point of bin n as $w^n, n = 1, 2, \ldots, N$, which represent the discretized scenarios of the uncertain parameter. Based on the definition of the Wasserstein metric and the construction of $\mathfrak{D}(\mathrm{P}, \hat{\mathrm{P}})$ in (2), we can reformulate the ambiguity set $\mathbb{D}$ in (2) by the following constraints:

$$\sum_{n=1}^{N} \sum_{m=1}^{N} \mathrm{q}^{nm} |w^m - w^n| \le \alpha, \tag{5}$$

$$\sum_{n=1}^{N} \mathrm{q}^{nm} = \mathrm{p}^m, \forall m \tag{6}$$

$$\sum_{m=1}^{N} \mathrm{q}^{nm} = \hat{\mathrm{p}}^n, \forall n \tag{7}$$

$$\sum_{m=1}^{N} \mathrm{p}^m = 1, \tag{8}$$

where p^n and $\hat{\mathrm{p}}^n$ represent the true probability and reference probability of scenario n respectively, and q^{nm}, $n = 1, \ldots N, m = 1, \ldots N$ denotes the joint probability distribution (i.e., $\mathbb{Q}$ in (2)). Constraint (5) is a reformulation of (2). Constraints (6) and (7) represent that $\mathrm{p}^m, m = 1, \ldots, N$ and $\hat{\mathrm{p}}^n, n = 1, \ldots, N$ are marginal distributions of q^{nm}, $n = 1, \ldots N, m = 1, \ldots N$.

III. Problem Formulation

In this section, we develop a data-driven stochastic defender-attacker-defender model for the transmission system hardening planning problem considering uncertain wind power generation and unknown disruptive events, such as natural disasters.

As a matter of fact, although natural disasters (or terrorist attacks) happen infrequently and the related historical records are quite limited, they can bring catastrophic impacts when they happen. However, using historical records, we can claim that the number of transmission lines disrupted simultaneously is no more than U. Therefore, we formulate the uncertainty set of random disruptions as follows:

$$\mathbb{U} = \left\{ \sum_{(i,j)\in\mathcal{E}} (1-u_{ij}) \le U, u_{ij} \in \{0,1\} \right\}. \tag{9}$$

In addition, natural disasters are intrinsically correlated in location. In order to consider this fact in our proposed model, we can restrict the random disruptions to the area that is vulnerable to natural disasters by adjusting the uncertainty set $\mathbb{U}$. In this case, we replace $\mathcal{E}$ in (9) with $\mathcal{E}_v$, which denotes the transmission lines in the vulnerable area (see [12]).

To hedge against the risk brought by natural disasters and wind energy intermittency to the power system, we consider the joint worst case of disruption scenario in $\mathbb{U}$ and probability distribution of wind output in $\mathbb{D}$. Accordingly, we develop a data-driven two-stage stochastic model that considers the hardening decision variables in the first stage, and deals with power generation level, power flow, phase angle and load shedding variables by considering the worst-case transmission line disruption in $\mathbb{U}$ and the worst-case probability distribution of wind output in $\mathbb{D}$ in the second stage. We formulate a data-driven two-stage stochastic transmission system hardening planning model as follows:

$$\min_{z} \sum_{(i,j)\in\mathcal{E}} H_{ij} z_{ij} + \max_{u\in\mathbb{U}} \max_{\mathrm{P}\in\mathbb{D}} E_{\mathrm{P}}[\mathcal{Q}(z,u,w(\xi))] \tag{10}$$

$$\text{s.t.} \quad z_{ij} \in \{0,1\}, \forall (i,j) \in \mathcal{E}, \tag{11}$$

where,

$$\mathcal{Q}(z,u,w(\xi)) = \min_{p,f,\theta,s} \sum_{i} \sum_{t} L_{it} s_{it}(\xi) \tag{12}$$

$$\begin{aligned} \text{s.t.} \quad & (z_{ij}+u_{ij}-z_{ij}u_{ij})(\theta_{it}(\xi)-\theta_{jt}(\xi)) \\ & - X_{ij} f_{ij,t}(\xi) = 0, \forall t \in \mathcal{T}, \forall (i,j) \in \mathcal{E}, \end{aligned} \tag{13}$$

$$\begin{aligned} & -F_{ij}(z_{ij}+u_{ij}-z_{ij}u_{ij}) \le f_{ij,t}(\xi) \\ & \le F_{ij}(z_{ij}+u_{ij}-z_{ij}u_{ij}), \forall t \in \mathcal{T}, \forall (i,j) \in \mathcal{E}, \end{aligned} \tag{14}$$

$$\begin{aligned} & p_{it}(\xi) + \sum_{j\in\mathcal{B}_i(.,i)} f_{ji,t}(\xi) - \sum_{j\in\mathcal{B}_i(i,.)} f_{ij,t}(\xi) + s_{it}(\xi) \\ & = d_{it} - w_{it}(\xi), \forall t \in \mathcal{T}, \forall i \in \mathcal{B}, \end{aligned} \tag{15}$$

$$p_{it}(\xi) \le C_i, \forall t \in \mathcal{T}, \forall i \in \mathcal{B}, \tag{16}$$

$$\theta_i^{\min} \le \theta_{it}(\xi) \le \theta_i^{\max}, \forall t \in \mathcal{T}, \forall i \in \mathcal{B}, \tag{17}$$

$$p_{it}(\xi), s_{it}(\xi) \ge 0, \forall t \in \mathcal{T}, \forall i \in \mathcal{B}, \tag{18}$$

where ξ represents randomness of wind output in the model. The objective function (10) is to minimize the system cost, i.e., hardening cost plus expected load shed cost under the worst-case disruption scenario and the worst-case distribution of wind output. Note here that in literature, hardening decisions or costs are often considered as budget constraints. In our paper, we include hardening costs in the objective function instead so that we can show different numbers of hardened lines led by different approaches. Constraints (13) represent the power flow in terms of phase angles. Constraints (14) are the transmission line flow capacity limits. In practice, hardening does not eliminate the chance of vulnerability to natural disasters. To address this, we can add a chance constraint to reflect the chance of vulnerability for hardened lines into the model. However, since hardening can reduce the chance of vulnerability significantly to a very small level, in this paper, we neglect the small chance of vulnerability of hardened lines and assume that the hardened lines are invulnerable to disruptions. Accordingly, the term $z_{ij}+u_{ij}-z_{ij}u_{ij}$ ensures that power flow constraints hold for any status of hardening and disruption. If line (i,j) is hardened ($z_{ij}=1$), then $z_{ij}+u_{ij}-z_{ij}u_{ij}=1$. So, this line is invulnerable and the power flow constraint holds. If line (i,j) is not hardened ($z_{ij,t}=0$), then $z_{ij}+u_{ij}-z_{ij}u_{ij}=u_{ij}$. Hence, the power flow depends on the attack scenario. If line (i,j) is attacked, i.e., $u_{ij}=0$, then there is no flow on line (i,j) and (13) and (14) hold. If line (i,j) is not attacked, i.e., $u_{ij}=1$, the power flow constraint also holds. Constraints (15) observe the power balance at each bus. Constraints (16) are the thermal generation capacity limits. Constraints (17) enforce the phase angle limit at each bus.

Note here that power flow constraints (13) and (14) are nonlinear. To linearize them, we fix the first-stage hardening decisions z_{ij} and consider them as input parameters in the second-stage problem. We consider two cases based on the hardening decisions derived from the first stage. In case one, we define set $\mathcal{E}_H = \{(i,j) \in \mathcal{E} | z_{ij}=1\}$ as the subset of hardened lines. The transmission lines in $\mathcal{E}_H$ are not affected by disruptions, and then for $\forall t \in \mathcal{T}, \forall (i,j) \in \mathcal{E}_H$, constraints (13) and (14) can be reformulated as:

$$(\theta_{it}(\xi)-\theta_{jt}(\xi)) - X_{ij} f_{ij,t}(\xi) = 0, \tag{19}$$

$$-F_{ij} \le f_{ij,t}(\xi) \le F_{ij}. \tag{20}$$

In case two, for the transmission lines that are not in set $\mathcal{E}_H$, i.e., the lines are not protected, the power flows in these lines depend on the disruption status. In this case, for $\forall t \in \mathcal{T}, \forall (i,j) \in \mathcal{E}\backslash\mathcal{E}_H$, we adopt the big-M method to linearize power flow constraints (13) and (14) as follows:

$$(\theta_{it}(\xi)-\theta_{jt}(\xi)) - X_{ij} f_{ij,t}(\xi) + M(1-u_{ij}) \ge 0, \tag{21}$$

$$(\theta_{it}(\xi)-\theta_{jt}(\xi)) - X_{ij} f_{ij,t}(\xi) - M(1-u_{ij}) \le 0, \tag{22}$$

$$-F_{ij}u_{ij} \le f_{ij,t}(\xi) \le F_{ij}u_{ij}. \tag{23}$$

Now, our data-driven model can be presented as the following program:

$$\min_{z} \sum_{(i,j)\in\mathcal{E}} H_{ij} z_{ij} + \max_{u\in\mathbb{U},\mathrm{P}\in\mathbb{D}} \sum_{m=1}^{N} \mathrm{p}_{it}^{m} \cdot \min_{p,f,\theta,s,o} \sum_{i}\sum_{t} L_{it} s_{it}(\xi^{m})$$
$$\text{s.t.} \quad \text{Constraints (11), (15)–(23).} \tag{24}$$

In the second-stage objective function, due to the independence of scenarios, we can interchange the second-stage minimization and summation (corresponding to the expectation term) operations. Then, we are able to reformulate our data-driven two-stage stochastic transmission system hardening planning problem as follows:

$$\min_{z} \sum_{(i,j)\in\mathcal{E}} H_{ij} z_{ij} + \max_{\mathrm{p},\mathrm{q},u} \min_{p,f,\theta,s,o} \sum_{m=1}^{N} \sum_{i}\sum_{t} \mathrm{p}_{it}^{m} L_{it} s_{it}(\xi^{m})$$
$$\text{s.t.} \quad \text{Constraints (5)–(9),}$$
$$\text{Constraints (11), (15)–(23).} \tag{25}$$

Notice here, in our two-stage model, hardening decisions are made in the first stage (defender's hardening plan). Second-stage is a max-min problem, in which the outer maximization problem considers the worst-case probability distribution of random wind output and the worst-case random disruption (attacker's plan) and inner the minimization problem minimizes the system cost by re-dispatching the power output (defender's corrective actions). To solve this model, we transform our two-stage data-driven formulation into a tractable reformulation and employ a decomposition algorithm. In the following section, we describe the approach to reformulate the model and solution algorithm in details.

IV. Solution Methodology

In this section, first, we explain how to reformulate (25) into a tractable reformulation. Second, we employ a decomposition algorithm to solve this reformulation. In order to reformulate the second-stage problem into a tractable formulation, we dualize the inner minimization problem and combine it to the outer maximization problem. Since the inner minimization is a linear program, which will lead to no dualization gap. The dual form of the inner minimization problem can be presented as follows:

$$\max_{\lambda,\gamma,\mu,\eta,\sigma,\tau,\upsilon} \sum_{m=1}^{N} \Bigg(\sum_{i}\sum_{t} \big((d_{it}^{m} - w_{it}^{m})\lambda_{it}^{m} - C_i\gamma_{it}^{m} + \theta_i^{\min}\bar{\eta}_{it}^{m} - \theta_i^{\max}\tilde{\eta}_{it}^{m}\big) - \sum_{(i,j)\in\mathcal{E}\setminus\mathcal{E}_H} \sum_{t} \big(F_{ij} u_{ij} (\bar{\tau}_{ij,t}^{m} + \tilde{\tau}_{ij,t}^{m}) + M(1-u_{ij})(\bar{\sigma}_{ij,t}^{m} + \tilde{\sigma}_{ij,t}^{m})\big) - \sum_{(i,j)\in\mathcal{E}_H} \sum_{t} \big(F_{ij}(\bar{\mu}_{ij,t}^{m} + \tilde{\mu}_{ij,t}^{m})\big) \Bigg) \tag{26}$$

s.t.

$$\lambda_{it}^{m} - \gamma_{it}^{m} \le 0,\ \forall t, \forall i, \forall m \tag{27}$$

$$\lambda_{it}^{m} - \lambda_{jt}^{m} + \bar{\tau}_{ij,t}^{m} - \tilde{\tau}_{ij,t}^{m} - X_{ij}\bar{\sigma}_{ij,t}^{m} + X_{ij}\tilde{\sigma}_{ij,t}^{m} = 0, \quad \forall (i,j) \in \mathcal{E}\setminus\mathcal{E}_H, \forall t, \forall m \tag{28}$$

$$\lambda_{it}^{m} - \lambda_{jt}^{m} + \bar{\mu}_{ij,t}^{m} - \tilde{\mu}_{ij,t}^{m} - X_{ij}\upsilon_{ij,t}^{m} = 0, \quad \forall (i,j) \in \mathcal{E}_H, \forall t, \forall m \tag{29}$$

$$\sum_{(i,j)\in\mathcal{E}\setminus\mathcal{E}_H(i,.)} (\bar{\sigma}_{ij,t}^{m} - \tilde{\sigma}_{ij,t}^{m}) + \sum_{(i,j)\in\mathcal{E}\setminus\mathcal{E}_H(.,i)} (\tilde{\sigma}_{ij,t}^{m} - \bar{\sigma}_{ij,t}^{m}) + \sum_{(i,j)\in\mathcal{E}_H(i,.)} \upsilon_{ij,t}^{m} - \sum_{(i,j)\in\mathcal{E}_H(.,i)} \upsilon_{ij,t}^{m} + \bar{\eta}_{it}^{m} - \tilde{\eta}_{it}^{m} = 0,\ \forall t, \forall i, \forall m \tag{30}$$

$$\lambda_{it}^{m} - L_{it}\mathrm{p}_{it}^{m} \le 0,\ \forall t, \forall i, \forall m \tag{31}$$

$$\bar{\eta}_{it}^{m}, \tilde{\eta}_{it}^{m}, \bar{\tau}_{it}^{m}, \tilde{\tau}_{it}^{m}, \bar{\mu}_{ij,t}^{m}, \tilde{\mu}_{ij,t}^{m}, \bar{\sigma}_{ij,t}^{m}, \tilde{\sigma}_{ij,t}^{m} \ge 0,$$
$$\gamma_{it}^{m}, \mathrm{p}_{it}^{m} \ge 0, \lambda_{it}^{m}, \upsilon_{it}^{m} \text{ are free}, \forall t, \forall i, \forall m \tag{32}$$

where λ_{it}^{m} and γ_{it}^{m} are dual variables of constraints (15) and (16), respectively; $\bar{\eta}_{it}^{m}$ and $\tilde{\eta}_{it}^{m}$ are dual variables of constraint (17); $\upsilon_{ij,t}^{m}$ are dual variables of constraint (19); $\bar{\mu}_{ij,t}^{m}$ and $\tilde{\mu}_{ij,t}^{m}$ are dual variables of constraint (20); $\bar{\sigma}_{ij,t}^{m}$ and $\tilde{\sigma}_{ij,t}^{m}$ are dual variables of constraints (21) and (22); $\bar{\tau}_{ij,t}^{m}$ and $\tilde{\tau}_{ij,t}^{m}$ are dual variables of constraint (23);

A. Linearizing the Objective Function

Notice there are bilinear terms in the objective function (26) of the derived reformulation, we use the well-known Big-M method to linearize bilinear terms. We let $\sum_m \sum_{(i,j)\in\mathcal{E}\setminus\mathcal{E}_H} \sum_t u_{ij}\bar{\tau}_{ij,t}^{m} = \sum_{(i,j)\in\mathcal{E}\setminus\mathcal{E}_H} \bar{\nu}_{ij}$ and $\sum_m \sum_{(i,j)\in\mathcal{E}\setminus\mathcal{E}_H} \sum_t u_{ij}\tilde{\tau}_{ij,t}^{m} = \sum_{(i,j)\in\mathcal{E}\setminus\mathcal{E}_H} \tilde{\nu}_{ij}$. Then, we have

$$\sum_{m,(i,j)\in\mathcal{E}\setminus\mathcal{E}_H,t} F_{ij} u_{ij} (\bar{\tau}_{ij,t}^{m} + \tilde{\tau}_{ij,t}^{m}) = \sum_{(i,j)\in\mathcal{E}\setminus\mathcal{E}_H} F_{ij}(\bar{\nu}_{ij} + \tilde{\nu}_{ij}) \tag{33}$$

$$\text{s.t.} \quad \bar{\nu}_{ij} \ge \sum_{m}\sum_{t} \bar{\tau}_{ij,t}^{m} - M(1-u_{ij}), \forall (i,j) \in \mathcal{E}\setminus\mathcal{E}_H, \tag{34}$$

$$\bar{\nu}_{ij} \ge -M u_{ij}, \forall (i,j) \in \mathcal{E}\setminus\mathcal{E}_H, \tag{35}$$

$$\tilde{\nu}_{ij} \ge \sum_{m}\sum_{t} \tilde{\tau}_{ij,t}^{m} - M(1-u_{ij}), \forall (i,j) \in \mathcal{E}\setminus\mathcal{E}_H, \tag{36}$$

$$\tilde{\nu}_{ij} \ge -M u_{ij}, \forall (i,j) \in \mathcal{E}\setminus\mathcal{E}_H. \tag{37}$$

In addition, we let $\sum_m \sum_{(i,j)\in\mathcal{E}\setminus\mathcal{E}_H} \sum_t u_{ij}\bar{\sigma}_{ij,t}^{m} = \sum_{(i,j)\in\mathcal{E}\setminus\mathcal{E}_H} \bar{\epsilon}_{ij}$ and $\sum_m \sum_{(i,j)\in\mathcal{E}\setminus\mathcal{E}_H} \sum_t u_{ij}\tilde{\sigma}_{ij,t}^{m} =$

$\sum_{(i,j)\in\mathcal{E}\backslash\mathcal{E}_H}\tilde{\epsilon}_{ij}$. Then, we have

$$\sum_{m,(i,j)\in\mathcal{E}\backslash\mathcal{E}_H,t} Mu_{ij}(\bar{\sigma}^m_{ij,t}+\tilde{\sigma}^m_{ij,t}) = \sum_{(i,j)\in\mathcal{E}\backslash\mathcal{E}_H} M(\bar{\epsilon}_{ij}+\tilde{\epsilon}_{ij}) \tag{38}$$

$$\text{s.t.}\quad \bar{\epsilon}_{ij} \le \sum_m\sum_t \bar{\sigma}^m_{ij,t} + M(1-u_{ij}), \forall (i,j)\in\mathcal{E}\backslash\mathcal{E}_H, \tag{39}$$

$$\bar{\epsilon}_{ij} \le Mu_{ij}, \forall (i,j)\in\mathcal{E}\backslash\mathcal{E}_H, \tag{40}$$

$$\tilde{\epsilon}_{ij} \le \sum_m\sum_t \tilde{\sigma}^m_{ij,t} + M(1-u_{ij}), \forall (i,j)\in\mathcal{E}\backslash\mathcal{E}_H, \tag{41}$$

$$\tilde{\epsilon}_{ij} \le Mu_{ij}, \forall (i,j)\in\mathcal{E}\backslash\mathcal{E}_H. \tag{42}$$

B. Decomposition Framework

In this paper, we employ the column-and-constraint generation approach [46] in a decomposition framework to solve the developed data-driven TSHP model. This approach is an iterative approach. In each iteration, the master problem (MP) is a relaxation of the original problem, which aims to find a lower bound of the problem and a current best hardening decision (may not be feasible). The sub-problem (SUB) is a reformulation of the second-stage problem, with the objective of obtaining an upper bound of the original problem and the worst-case scenarios. This iterative procedure continues until the difference between the upper bound and the lower bound is no more than a predefined level. As a result, the optimal hardening planning decision will be output and the worst-case wind output distribution and disruption scenario will be identified.

In our model, since the first-stage decisions are hardening decisions, they do not affect the feasibility of the second-stage problem. Moreover, since we allow load shedding in the model, the second-stage problem is always feasible for any hardening decision z derived from the first stage. Hence, there is no need for the first-stage decision feasibility check. Then, we have the following MP.

$$\begin{aligned}\min_{z}\quad & \sum_{(i,j)\in\mathcal{E}} H_{ij}z_{ij} + \theta\\ \text{s.t. Constraints (11)},&\\ \text{Constraints (15)–(18), for each } \xi^m&\\ \text{Optimality cuts},&\end{aligned}$$

where θ is set to be the second-stage objective value. Based on the reformulation of (33) to (42), we can reformulate the sub-problem as below:

$$\begin{aligned}\psi(z) = \max_{\lambda,\gamma,\mu,\eta,\sigma,\tau,\epsilon,\upsilon,\nu,\mathrm{p},\mathrm{q},u} & \sum_{m=1}^{N}\\ \Bigg(\sum_i\sum_t & \left((d^m_{it}-w^m_{it})\lambda^m_{it} - C_i\gamma^m_{it} + \theta^{\min}_i\bar{\eta}^m_{it} - \theta^{\max}_i\tilde{\eta}^m_{it}\right)\\ - \sum_{(i,j)\in\mathcal{E}\backslash\mathcal{E}_H}\sum_t & \left(M(\bar{\sigma}^m_{ij,t}+\tilde{\sigma}^m_{ij,t})\right)\\ - \sum_{(i,j)\in\mathcal{E}\backslash\mathcal{E}_H}\sum_t & F_{ij}(\bar{\mu}^m_{ij,t}+\tilde{\mu}^m_{ij,t})\Bigg)\\ + \sum_{(i,j)\in\mathcal{E}\backslash\mathcal{E}_H} & \left(M(\bar{\epsilon}_{ij}+\tilde{\epsilon}_{ij}) - F_{ij}(\bar{\nu}_{ij}+\tilde{\nu}_{ij})\right)\\ & \text{Constraints (5)–(9), (27)–(32),}\\ & \text{Constraints (34)–(37), (39)–(42).}\end{aligned} \tag{43}$$

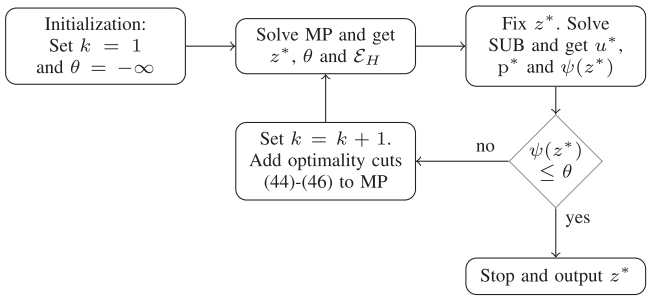


Fig. 1. Flowchart of the column-and-constraint generation algorithm.

C. Column-and-Constraint Generation Algorithm

The column-and-constraint generation algorithm is summarized in the following steps and in the flowchart presented in Fig. 1.:

1. Initialization. Set $k = 1, \theta = -\infty$.
2. Solve MP and get the first-stage optimal solution (the optimal hardening plan) z^*, θ and $\mathcal{E}_H$.
3. Solve SUB for the current hardening plan z^* to obtain the worst-case disruption scenario u^*, the worst-case wind output distribution p^* and $\psi(z^*)$.
4. If $\psi(z^*) \le \theta$, stop. Current z^* is the optimal hardening plan. Output the result. Otherwise, set $k = k+1$. For the disruption plan u^*, let subset $\mathcal{E}_A = \{(i,j)\in\mathcal{E} | u_{ij} = 0\}$ be the set of attacked transmission lines. Then, add the following constraints to MP and go to step 2.

$$\theta \ge \sum_{m=1}^{N}\sum_i\sum_t \mathrm{p}^{m*}_{it}L_{it}s_{it}(\xi^m), \tag{44}$$

$$\begin{aligned}&(\theta_{it}(\xi^m) - \theta_{jt}(\xi^m)) - X_{ij}f_{ij,t}(\xi^m) + M(1-z_{ij}) \ge 0,\\ &(\theta_{it}(\xi^m) - \theta_{jt}(\xi^m)) - X_{ij}f_{ij,t}(\xi^m) - M(1-z_{ij}) \le 0,\\ &-F_{ij}z_{ij} \le f_{ij,t}(\xi^m) \le F_{ij}z_{ij},\\ &\forall t, \forall m, \forall (i,j)\in\mathcal{E}_A,\end{aligned} \tag{45}$$

$$\begin{aligned}&(\theta_{it}(\xi^m) - \theta_{jt}(\xi^m)) - X_{ij}f_{ij,t}(\xi^m) = 0,\\ &-F_{ij} \le f_{ij,t}(\xi^m) \le F_{ij},\\ &\forall t, \forall m, \forall (i,j)\in\mathcal{E}\backslash\mathcal{E}_A.\end{aligned} \tag{46}$$

V. Numerical Results

In this section, we conduct numerical experiments to test the performance of the proposed approach. We apply our approach to a 24-node system [47], which is based on the IEEE one-area RTS-96 test system [48], and a modified IEEE 118-bus test

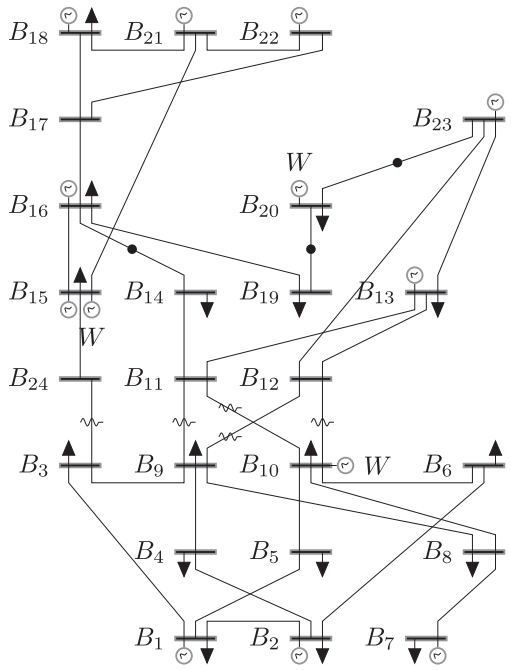


Fig. 2. Modified 24-node system and the DDTSHP hardening plan for $U = 4$.

system (available at http://motor.ece.iit.edu/data). According to [49], wind energy constitutes 5.6% of the total electricity generation across the united states in 2016. Accordingly, we modify both test systems by adding several wind power generation capacities, which account for 10% of the total generation capacity. We compare the system performance between the proposed data-driven transmission system hardening planning (DDTSHP) and Robust transmission system hardening planning (ROTSHP). We also discuss how the historical data and the ambiguity set can affect the conservativeness of the proposed data-driven approach. To implement the proposed algorithm, we use C++ and CPLEX 12.6 and run it on a computer with Intel(R) Xeon(R) 3.2 GHz and 8 GB memory.

A. Data Generation

In order to generate the set of historical data, we use a Monte Carlo simulation. For simulation convenience, we assume that wind outputs are independent for different load blocks. Also, for each load block, we assume that the unknown wind output follows a normal distribution with the forecasted wind output as the mean and 0.3 of the mean value as the standard deviation. To generate wind output scenarios, we generate samples for each wind farm and each load block and set the number of bins to be 5.

B. 24-Node System

This system, depicted in Fig. 2, consists of 24 buses, 12 generators, 34 transmission lines and 17 loads. We consider all generators to be thermal plants and add three wind farms at buses 10, 15 and 20. The first column of Table I shows the index of transmission lines and the second and third columns show origins and ends of lines respectively.

First, we compare the proposed DDTSHP with ROTSHP. We numerically show the DDTSHP obtains less conservative hardening plans compared to ROTSHP. To this end, we let the attack budget vary within a range of 1 to 10. We set the size of historical data set S and the confidence level β to be 100 and 0.99%, respectively. With a set of generated data in Section V-A, we solve the modified 24-node test system for different attack budgets and get the optimal first-stage hardening plans using both DDTSHP and ROTSHP. Then, we fix the optimal hardening decisions and solve the second-stage problem with a new set of randomly simulated wind output and attack scenarios for different attack budgets. In Table II, we report the optimal hardening plans obtained by DDTSHP and ROTSHP for various attack budgets. In Table III, we compare the number of hardened transmission lines (denoted by NHL), total cost (denoted by Obj), and the computational time in seconds (denoted by CPU). Notice that in Table III, there is no load shed for attack budget $U = 1$. However, as the attack budget U increases, the total cost increases for both DDTSHP and ROTSHP. We can also observe that DDTSHP results in significant cost reduction compared to that of ROTSHP, especially when $U = 4$, DDTSHP reduces the cost by 30.5% (the optimal hardening plan identified

TABLE I
TRANSMISSION LINES OF 24-NODE SYSTEM

No.	From	To	No.	From	To	No.	From	To
1	1	2	13	8	10	25	15	21
2	1	3	14	9	11	26	15	24
3	1	5	15	9	12	27	16	17
4	2	4	16	10	11	28	16	19
5	2	6	17	10	12	29	17	18
6	3	9	18	11	13	30	17	22
7	3	24	19	11	14	31	18	21
8	4	9	20	12	13	32	19	20
9	5	10	21	12	23	33	20	23
10	6	10	22	13	23	34	21	22
11	7	8	23	14	16			
12	8	9	24	15	16			

TABLE II
HARDENING PLANS OF DDTSHP VERSUS ROTSHP FOR 24-NODE SYSTEM

Attack Budget	Hardening Plan	
	DDTSHP	ROTSHP
2	(5, 23, 28)	(5, 23, 32, 33)
3	(6, 32, 33)	(11, 19, 32, 33)
4	(23, 32, 33)	(11, 18, 22, 32, 33)
5	(6, 11, 14, 18, 32, 33)	(11, 14, 18, 22, 32, 33)
6	(6, 11, 14, 18, 32, 33)	(6, 11, 14, 16, 18, 32, 33)
7	(6, 11, 14, 18, 32, 33)	(6, 11, 14, 16, 18, 32, 33)
8	(6, 11, 14, 18, 32, 33)	(6, 10, 11, 14, 17, 18, 21, 32, 33)
9	(6, 11, 14, 18, 32, 33)	(6, 10, 11, 14, 17, 18, 21, 32, 33)
10	(6, 11, 14, 18, 32, 33)	(6, 10, 11, 14, 17, 18, 21, 32, 33)

TABLE III
PERFORMANCE OF DDTSHP VERSUS ROTSHP FOR 24-NODE SYSTEM

Attack Budget	DDTSHP			ROTSHP		
	NHL	Obj ($m)	CPU(s)	NHL	Obj ($m)	CPU(s)
1	0	0	0.8	0	0	0.8
2	3	2.6209	8.5	4	2.8865	5.5
3	3	2.6465	18.8	4	3.4887	10.9
4	3	3.1073	33.1	5	4.4717	15.4
5	6	5.0757	30.6	6	5.3680	13.8
6	6	5.4582	27.6	7	6.0558	15.3
7	6	5.5317	28.1	7	6.4745	10.8
8	6	5.9142	21.0	9	6.6830	10.3
9	6	6.0783	24.2	9	6.6925	12.1
10	6	6.1923	23.0	9	6.7198	11.6

TABLE IV
EFFECTS OF THE HISTORICAL DATA ON THE LOAD SHEDDING COST ($M)

Attack Budget	# of data					
	50	100	500	1000	5000	10000
2	2.6256	2.6209	2.6126	2.6114	2.6105	2.6104
3	2.6480	2.6464	2.6448	2.6445	2.6443	2.6442
4	3.4665	3.4595	3.4486	3.4470	3.4458	3.4457
5	5.7041	5.6919	5.6808	5.6794	5.6782	5.6780
6	5.9675	5.9541	5.9444	5.9430	5.9417	5.9415
7	6.0234	6.0122	6.0031	6.0012	5.9998	5.9996
8	6.2500	6.2482	6.2292	6.2276	6.2265	6.2350
9	6.5043	6.4928	6.4825	6.4810	6.4798	6.4796
10	6.5745	6.5675	6.5533	6.5523	6.5507	6.5505

by DDTSHP (asterisked lines) for $U = 4$ is presented in Fig. 2). In addition, we see for most attack budgets that DDTSHP's hardening plans include a lesser number of hardened lines than those of ROTSHP. From these results, we can claim that the proposed DDTSHP approach is less conservative than ROTSHP. That is because DDTSHP uses the data information and considers the worst-case distribution of the wind output within the ambiguity set, while ROTSHP considers the worst-case wind output scenario. Hence, ROTSHP leads to a higher number of transmission lines to be hardened and therefore, this over-conservativeness results in higher hardening costs and accordingly higher total costs. Furthermore, DDTSHP leads to more computational efforts than ROTSHP due to the consideration of the stochastic nature of wind output, so there is a tradeoff between the cost reduction and computational efficiency by using DDTSHP.

Second, we numerically illustrate and discuss how the size of the historical data set can affect the conservativeness of the proposed DDTSHP. Accordingly, we conduct numerical experiments on the modified 24-node test system for different attack budgets from 2 to 10 and represent the results in Table IV. Here, we allow the size of historical data to vary between 50 to 10000 and set the confidence level β to be 99%. As shown in Table IV, as the size of historical data increases, the total cost decreases. That is because, according to equation (4), there is an inverse correlation between the value of α and the size of historical data set S. As the number of historical observations increases, the value of α decreases; and hence, the confidence set $\mathbb{D}$ shrinks. From the theoretical point of view, with an infinite number of historical observations, the value of α eventually goes to zero. In fact, in this case, the reference distribution converges to the true distribution and the proposed data-driven approach becomes risk neutral. Moreover, we can see that larger attack budgets lead to higher total costs.

Third, we assess the effect of the ambiguity set $\mathbb{D}$ on the performance of the proposed data-driven approach. Here, we set the size of historical data S to be 100 and test our DDTSHP on the modified 24-node test system. Note here that, according to (4), as the value of β increases, the value of α increases. Thus, a larger value of β leads to a bigger ambiguity set $\mathbb{D}$. Therefore, we can control the size of ambiguity set by adjusting the confidence level β. We allow the confidence level β to vary within a range from 0.5 to 0.99 and show the associated system costs for different values of attack budget in Table V. From the results shown in Table V we can see that as the value of β increases, our proposed data-driven approach becomes more conservative and the total system cost increases.

TABLE V
EFFECTS OF THE AMBIGUITY SET ON THE LOAD SHEDDING COST ($M)

Attack Budget	Confidence level β					
	0.5	0.7	0.8	0.9	0.95	0.99
2	2.6170	2.6173	2.6175	2.6182	2.6188	2.6209
3	2.6453	2.6456	2.6457	2.6458	2.6459	2.6464
4	3.4506	3.4525	3.4547	3.4553	3.4552	3.4595
5	5.6841	5.6852	5.6861	5.6871	5.6880	5.6919
6	5.9474	5.9476	5.9481	5.9499	5.9508	5.9541
7	6.0045	6.0057	6.0070	6.0079	6.0107	6.0122
8	6.2312	6.2325	6.2328	6.2343	6.2354	6.2482
9	6.4847	6.4867	6.4879	6.4880	6.4889	6.4928
10	6.5557	6.5567	6.5574	6.5593	6.5613	6.5675

TABLE VI
DDTSHP VERSUS ROTSHP FOR 118-BUS SYSTEM

Attack Budget	DDTSHP			ROTSHP		
	NHL	Obj ($m)	CPU(s)	NHL	Obj ($m)	CPU(s)
2	4	1.8855	24.3	4	1.8855	13.8
3	4	1.9014	79.2	4	2.2706	30.2
4	4	2.2177	95.3	7	3.0744	42.8
5	6	3.0061	226.2	8	3.7661	78.7
6	6	3.2099	291.5	8	3.8470	78.4
7	6	3.4583	272.0	8	4.0783	115.9
8	6	3.6620	442.9	8	4.2694	105.0
9	6	3.9104	489.9	8	4.7832	90.6
10	10	5.1314	529.1	12	5.5279	98.7

C. 118-Bus System

In this section, we illustrate the effectiveness of our proposed DDTSHP approach on a larger system. The 118-bus test system consists of 118 buses, 33 generators and 186 transmission lines. We consider all generators to be thermal plants and add five wind farms at buses 20, 40, 60, 80 and 100. We consider a range of 2 to 10 for the attack budget. We also let the size of historical data S be 100 and the confidence level β be 99%. Then, we follow the same simulation procedure as the one in the previous subsection and compare the performance of DDTSHP with ROTSHP. We report the results in Table VI, which are consistent with the ones of 24-node system. That is, as the attack budget increases, the total system cost increases too. Moreover, as compared to ROTSHP, we see that DDTSHP leads to a lesser number of hardened lines and lower total costs, for most attack budgets. These results further show that DDTSHP is less conservative than ROTSHP.

VI. CONCLUSION

In this study, we developed an approach to deal with the stochastic transmission system hardening planning problem in the presence of wind generation uncertainty and multiple simultaneous disruptive events. Motivated by the considerable amount of historical data available to power system operators, we proposed a data-driven approach that learns from the wind output historical data and employs the Wasserstein metric to construct an ambiguity set for the unknown wind output probability distribution. Our data-driven two-stage model can obtain robust hardening decisions by considering the joint worst-case disruptive scenario and wind output distribution. We used a decomposition framework based on the Column-and-Constraints generation method to solve the proposed model. We showed through numerical experiments that although our approach is risk-averse, it leads to less conservative hardening decisions

than the robust optimization approach. In addition, we showed that the conservativeness of our approach depends on the number of available historical data and the confidence level we prefer.

APPENDIX
PROOF OF PROPOSITION 1

In order to prove Proposition 1, motivated by the approaches in nonparametric estimation of densities and distribution functions (see, e.g., [50]), we prove the following two lemmas first.

Lemma 1: Given a set of historical data of size S and N bins, we have

$$Pr(|\mathrm{p}^n - \hat{\mathrm{p}}^n| \geq \delta) \leq 2\exp(-2S\delta), \forall n = 1, \ldots, N. \tag{47}$$

Proof: Let y_s^n denote whether the observation s falls in the bin n, then $y_s^n \sim \text{Bernoulli}(\mathrm{p}^n), \forall n = 1, \ldots, N, \forall s = 1, \ldots, S$. We also have $\hat{\mathrm{p}}^n = \sum_{s=1}^{S} y_s^n / S$ and $E[\hat{\mathrm{p}}^n] = \mathrm{p}^n$. Then, according to Hoeffding's Inequality for Bernoulli random variables, (47) holds. ■

In order to get the convergence rate under the Wasserstein metric, we will first study another metric, i.e., the Total Variance metric, and by taking advantage of the relationship between the two metrics, we will obtain the convergence rate of the Wasserstein metric. According to [51], for space $\mathcal{W}$, the Total Variation metric is defined as $\mathfrak{D}_{\mathrm{TV}}(\mathrm{P}, \hat{\mathrm{P}}) = \sum_{n=1}^{N} |\mathrm{p}^n - \hat{\mathrm{p}}^n|$.

Lemma 2: Given a set of historical data of size S and N bins, under the Total Variation metric we have

$$Pr(\mathfrak{D}_{\mathrm{TV}}(\mathrm{P}, \hat{\mathrm{P}}) \geq \delta) \leq 2N\exp(-2S\delta/N). \tag{48}$$

Proof:

$$\begin{aligned} Pr(\mathfrak{D}_{\mathrm{TV}}(\mathrm{P}, \hat{\mathrm{P}}) \geq \delta) &= Pr\left(\sum_{n=1}^{N} |\mathrm{p}^n - \hat{\mathrm{p}}^n| \geq \delta\right) \\ &\leq Pr(\cup_{n=1}^{N} [|\mathrm{p}^n - \hat{\mathrm{p}}^n| \geq \delta/N]) \\ &\leq \sum_{n=1}^{N} Pr([|\mathrm{p}^n - \hat{\mathrm{p}}^n| \geq \delta/N]) \\ &\leq 2N\exp(-2S\delta/N). \end{aligned} \tag{49}$$

where the inequality (49) holds due to Lemma 1. ■

Based on the above two lemmas, we are ready to prove Proposition 1. First, according to [51], we have the following relationship between the Wasserstein metric and the Total Variation metric:

$$\mathfrak{D}_{\mathrm{w}}(\mathrm{P}, \hat{\mathrm{P}}) \leq \frac{D}{2} \quad \mathfrak{D}_{\mathrm{TV}}(\mathrm{P}, \hat{\mathrm{P}}), \tag{50}$$

where D denoted the diameter of the supporting space $\mathcal{W}$. Moreover, according to Lemma 2, we have

$$Pr(\mathfrak{D}_{\mathrm{TV}}(\mathrm{P}, \hat{\mathrm{P}}) \leq \delta) \geq 1 - 2N\exp(-2S\delta/N). \tag{51}$$

Inequality (50) implies that $\frac{2\mathfrak{D}_{\mathrm{w}}(\mathrm{P}, \hat{\mathrm{P}})}{D} \leq \mathfrak{D}_{\mathrm{TV}}(\mathrm{P}, \hat{\mathrm{P}})$. Therefore,

$$\begin{aligned} Pr\left(\frac{2\mathfrak{D}_{\mathrm{w}}(\mathrm{P}, \hat{\mathrm{P}})}{D} \leq \delta\right) &\geq Pr(\mathfrak{D}_{\mathrm{TV}}(\mathrm{P}, \hat{\mathrm{P}}) \leq \delta) \\ &\geq 1 - 2N\exp(-2S\delta/N). \end{aligned} \tag{52}$$

Let $\alpha = D\delta/2$, then the proof of Proposition 1 is done.

REFERENCES

[1] A. B. Smith and R. W. Katz, "US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases," *Natural Hazards*, vol. 67, no. 2, pp. 387–410, Feb. 2013.

[2] Executive Office of the President, "Economic benefits of increasing electric grid resilience to weather outages," August 2013 [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6978705

[3] J. Salmeron, K. Wood, and R. Baldick, "Analysis of electric grid security under terrorist threat," *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 905–912, May 2004.

[4] N. Romero, L. K. Nozick, I. Dobson, N. Xu, and D. A. Jones, "Seismic retrofit for electric power systems," *Earthquake Spectra*, vol. 31, no. 2, pp. 1157–1176, 2015.

[5] A. L. Motto, J. M. Arroyo, and F. D. Galiana, "A mixed-integer LP procedure for the analysis of electric grid security under disruptive threat," *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1357–1365, Aug. 2005.

[6] J. M. Arroyo, "Bilevel programming applied to power system vulnerability analysis under multiple contingencies," *IET Gener. Transm. Distrib.*, vol. 4, no. 2, pp. 178–190, Feb. 2010.

[7] Y. Yao, T. Edmunds, D. Papageorgiou, and R. Alvarez, "Trilevel optimization in power network defense," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 4, pp. 712–718, Jul. 2007.

[8] G. Brown, M. Carlyle, J. Salmerón, and K. Wood, "Analyzing the vulnerability of critical infrastructure to attack and planning defenses," *Tuts. Operat. Res.: Emerg. Theory, Methods, Appl.*, pp. 102–123, 2005.

[9] V. M. Bier, E. R. Gratz, N. J. Haphuriwat, W. Magua, and K. R. Wierzbicki, "Methodology for identifying near-optimal interdiction strategies for a power transmission system," *Reliab. Eng. Syst. Saf.*, vol. 92, no. 9, pp. 1155–1161, Sep. 2007.

[10] W. Yuan, L. Zhao, and B. Zeng, "Optimal power grid protection through a defender–attacker–defender model," *Rel. Eng. Syst. Saf.*, vol. 121, pp. 83–89, Jan. 2014.

[11] N. Alguacil, A. Delgadillo, and J. M. Arroyo, "A trilevel programming approach for electric grid defense planning," *Comput. Oper. Res.*, vol. 41, pp. 282–290, Jan. 2014.

[12] W. Yuan, J. Wang, F. Qiu, C. Chen, C. Kang, and B. Zeng, "Robust optimization-based resilient distribution network planning against natural disasters," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2817–2826, Nov. 2016.

[13] F. Qiu, J. Wang, C. Chen, and J. Tong, "Optimal black start resource allocation," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 2493–2494, May 2016.

[14] M. Vaiman *et al.*, "Risk assessment of cascading outages: Methodologies and challenges," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 631–641, May 2012.

[15] N. Fan, D. Izraelevitz, F. Pan, P. M. Pardalos, and J. Wang, "A mixed integer programming approach for optimal power grid intentional islanding," *Energy Syst.*, vol. 3, no. 1, pp. 77–93, 2012.

[16] W. Sun, C.-C. Liu, and L. Zhang, "Optimal generator start-up strategy for bulk power system restoration," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1357–1366, Aug. 2011.

[17] F. Qiu and P. Li, "An integrated approach for power system restoration planning," *Proc. IEEE*, vol. 105, no. 7, pp. 1234–1252, Jul. 2017.

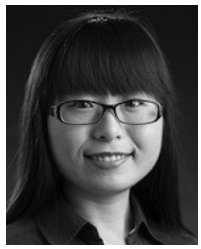
[18] M. Adibi and L. Fink, "Power system restoration planning," *IEEE Trans. Power Syst.*, vol. 9, no. 1, pp. 22–28, Feb. 1994.

[19] D. Lindenmeyer, H. Dommel, and M. Adibi, "Power system restoration-a bibliographical survey," *Int. J. Elec. Power*, vol. 23, no. 3, pp. 219–227, 2001.

[20] A. D. González, L. Dueñas-Osorio, M. Sánchez-Silva, and A. L. Medaglia, "The interdependent network design problem for optimal infrastructure system restoration," *Comput.-Aided Civil Inf.*, vol. 31, no. 5, pp. 334–350, 2016.

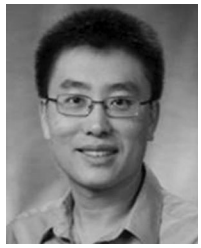
[21] L. Yutian, F. Rui, and V. Terzija, "Power system restoration: a literature review from 2006 to 2016," *J. Modern Power Syst. Clean*, vol. 4, no. 3, pp. 332–341, 2016.
[22] U.S. Department of Energy, "Hardening and resiliency: U.S. energy industry response to recent hurricane seasons," Aug. 2010. [Online]. Available: http://www.oe.netl.doe.gov/docs/HR-Re- port-final-081710.pdf
[23] Y. Wang, C. Chen, J. Wang, and R. Baldick, "Research on resilience of power systems under natural disasters-A review," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1604–1613, Mar. 2016.
[24] P. Kundur *et al.*, "Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1387–1401, Aug. 2004.
[25] K. W. Hedman, M. C. Ferris, R. P. O'Neill, E. B. Fisher, and S. S. Oren, "Co-optimization of generation unit commitment and transmission switching with N-1 reliability," *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 1052–1063, May 2010.
[26] North Electric Reliability Corporation, "NERC operating manual," Aug. 2016. [Online]. Available: http://www.nerc.com/comm/OC/Operating% 20Manual%20DL/Operating_Manual_20160809.pdf
[27] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation, and Control*. Hoboken, NJ, USA: Wiley, 2012.
[28] U.S. Department of Energy, "20% wind energy by 2030: Increasing wind energy's contribution to U.S. electricity supply," 2008. [Online]. Available: http://eere.energy.gov/wind/pdfs/41869.pdf
[29] F. Bouffard and F. D. Galiana, "Stochastic security for operations planning with significant wind power generation," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 306–316, May 2008.
[30] B. C. Ummels, M. Gibescu, E. Pelgrum, W. L. Kling, and A. J. Brand, "Impacts of wind power on thermal generation unit commitment and dispatch," *IEEE Trans. Energy Conver.*, vol. 22, no. 1, pp. 44–51, Mar. 2007.
[31] J. Wang, A. Botterud, V. Miranda, C. Monteiro, and G. Sheble, "Impact of wind power forecasting on unit commitment and dispatch," in *Proc. 8th Int. Workshop Large-Scale Integr. Wind Power Into Power Syst.*, Oct. 2009, pp. 1–8.
[32] P. A. Ruiz, C. R. Philbrick, E. Zak, K. W. Cheung, and P. W. Sauer, "Uncertainty management in the unit commitment problem," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 642–651, May 2009.
[33] C. Zhao and Y. Guan, "Unified stochastic and robust unit commitment," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3353–3361, Aug. 2013.
[34] C. Zhao, J. Wang, J.-P. Watson, and Y. Guan, "Multi-stage robust unit commitment considering wind and demand response uncertainties," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2708–2717, Aug. 2013.
[35] R. A. Jabr, "Robust transmission network expansion planning with uncertain renewable generation and loads," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4558–4567, Nov. 2013.
[36] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Oper. Res.*, vol. 58, no. 3, pp. 595–612, Jan. 2010.
[37] J. Goh and M. Sim, "Distributionally robust optimization and its tractable approximations," *Oper. Res.*, vol. 58, no. 4, pp. 902–917, Jul. 2010.
[38] R. Jiang and Y. Guan, "Data-driven chance constrained stochastic program," *Math. Program.*, vol. 158, no. 1/2, pp. 291–327, Jul. 2016.
[39] C. Zhao and Y. Guan, "Data-driven stochastic unit commitment for integrating wind generation," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2587–2596, Jul. 2016.
[40] W. Wei, F. Liu, and S. Mei, "Distributionally robust co-optimization of energy and reserve dispatch," *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 289–300, Jan. 2016.
[41] Q. Bian, H. Xin, Z. Wang, D. Gan, and K. P. Wong, "Distributionally robust solution to the reserve scheduling problem with partial information of wind power," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2822–2823, Sep. 2015.
[42] Y. Zhang, S. Shen, and J. L. Mathieu, "Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1378–1388, Mar. 2017.
[43] G. Pflug and D. Wozabal, "Ambiguity in portfolio selection," *Quant. Finance*, vol. 7, no. 4, pp. 435–442, Aug. 2007.
[44] D. Wozabal, "A framework for optimization under ambiguity," *Ann. Oper. Res.*, vol. 193, no. 1, pp. 21–47, Mar. 2012.
[45] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems: Volume I: Theory*. New York, NY, USA: Springer Science & Business Media, 1998, vol. 1.
[46] B. Zeng and L. Zhao, "Solving two-stage robust optimization problems using a column-and-constraint generation method," *Oper. Res. Lett.*, vol. 41, no. 5, pp. 457–461, Sep. 2013.
[47] A. J. Conejo, M. Carrión, and J. M. Morales, *Decision Making Under Uncertainty in Electricity Markets*. New York, NY, USA: Springer, 2010, vol. 1.
[48] C. Grigg *et al.*, "The IEEE reliability test system-1996. a report prepared by the reliability test system task force of the application of probability methods subcommittee," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 1010–1020, Aug. 1999.
[49] U.S. Energy Information Administration, "Electric power monthly," Oct. 2017. [Online]. Available: https://www.eia.gov/electricity/monthly/current_month/epm.pdf
[50] L. Devroye, "Exponential inequalities in nonparametric estimation," in *Proc. Nonparametric Functional Estimation Related Topics*. New York, NY, USA: Springer, 1991, pp. 31–44.
[51] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," *Int. Statist. Rev.*, vol. 70, no. 3, pp. 419–435, Dec. 2002.

Ali Bagheri (S'16) received the B.S. degree in applied mathematics from the Yazd University, Yazd, Iran, in 2005, and the M.S. degree in industrial engineering from Mazandaran University of Science and Technology, Babol, Iran, in 2008. He is currently working toward the Ph.D. degree with the Department of Industrial Engineering and Management, Oklahoma State University, Stillwater, OK, USA. His research interests include data-driven optimization, power system operations under uncertainty, and power system security.



Chaoyue Zhao (M'12) received the B.S. degree in information and computing sciences from the Fudan University, Shanghai, China, in 2010, and the Ph.D. degree in industrial and systems engineering from the University of Florida, Gainsville, FL, USA, in 2014. She is an Assistant Professor in industrial engineering and management with Oklahoma State University, Stillwater, OK, USA. Her research interests include data-driven stochastic optimization and stochastic integer programming with their applications in power grid security and renewable energy management. She worked at Pacific Gas and Electric Company in 2013.

Feng Qiu (M'14) received the Ph.D. degree from the School of Industrial and Systems Engineering, the Georgia Institute of Technology, Atlanta, Georgia, in 2013. He is a computational scientist with the Energy Systems Division at Argonne National Laboratory, Argonne, IL, USA. His current research interests include optimization in power system operations, electricity markets, and power grid resilience.



Jianhui Wang (M'07–SM'12) received the Ph.D. degree in electrical engineering from Illinois Institute of Technology, Chicago, Illinois, USA, in 2007. He is currently an Associate Professor with the Department of Electrical Engineering, Southern Methodist University (SMU), Dallas, Texas, USA. Prior to joining SMU, he had an eleven-year stint at Argonne National Laboratory with the last appointment as Section Lead – Advanced Grid Modeling. He is the secretary of the IEEE Power & Energy Society (PES) Power System Operations, Planning & Economics Committee. He has held visiting positions in Europe, Australia, and Hong Kong including a VELUX Visiting Professorship at the Technical University of Denmark. He is the Editor-in-Chief of the IEEE TRANSACTIONS ON SMART GRID and an IEEE PES Distinguished Lecturer. He is also the recipient of the IEEE PES Power System Operation Committee Prize Paper Award in 2015.