Boundary Defense against Cyber Threat for Power System Operation

Ming Jin* Javad Lavaei† Somayeh Sojoudi‡ Ross Baldick §

Abstract

The operation of power grids is becoming increasingly data-centric. While the abundance of data could improve the efficiency of systems, it poses major reliability challenges. In particular, state estimation aims to find the operating state of a network from the data, but an undetected attack on the data could lead to making wrong operational decisions for the system and trigger a large-scale blackout. Nevertheless, the understanding of the vulnerability of state estimation with regards to cyberattacks has been hindered by the lack of tools for studying the topological and data-analytic aspects of networks. Algorithmic robustness is critical in extracting reliable information from abundant but untrusted grid data. For a large-scale power grid, we quantify, analyze, and visualize the regions of the network that are not robust to cyberattacks in the sense that there exists a data manipulation strategy for each of those local regions that misleads the operator at the global scale and yields a wrong estimation of the state of the network at almost all buses. We also propose an optimization-based graphical boundary defense mechanism to identify the border of the geographical area with data that have been manipulated. The proposed method does not allow a local attack to have a global effect on the data analysis of the entire network, which enhances the situational awareness of the grid, especially in the face of adversity. The developed mathematical framework reveals key geometric and algebraic factors that can affect algorithmic robustness and is used to study the vulnerability of the U.S. power grid in this paper.

*Department of Industrial Engineering and Operation Research, University of California Berkeley, CA 94720, USA
†Department of Industrial Engineering and Operation Research, University of California Berkeley, CA 94720, USA
‡Department of Electrical Engineering and Computer Sciences, University of California Berkeley, CA 94720, USA
§Department of Electrical and Computer Engineering, University of Texas at Austin, TX 78712, USA

While real-world data abound for many complex systems, they are often noisy and corrupted. Acquiring reliable information from abundant but untrusted data is key to enhancing cybersecurity for mission-critical systems such as power grid [1]. Since many of these systems are inherently network structured, data analytics cannot be satisfactorily understood without incorporating their underlying graph topologies. For instance, consider the power system state estimation (SE) problem. SE monitors the operating status of the grid constantly by filtering and fusing a large volume of data every few minutes [2]. It plays a critical role in the economic and reliable operation of the grid because major operational problems, such as security-constrained optimal power flow, contingency analysis, and transient stability analysis, rely on its output. The current industry practice is based on a set of heuristic iterative algorithms proposed in the 70s, which are known empirically to work properly under normal situations. However, those algorithms become brittle under adverse conditions, such as natural hazards, equipment faults, and even cyberattacks, in which case part of the data is missing or manipulated maliciously. The significance of functioning SE was illustrated by the 2003 large-scale blackout, in which the failure of SE contributed to the inability of operators to provide real-time diagnostic support [3]. Despite substantial advances in algorithm design [2, 4–24], a major obstacle still remains: the lack of a framework for the design of a robust and scalable algorithm together with a realistic evaluation of its vulnerability. Developing such a framework is challenging for three reasons: (a) the model of a power system is highly nonlinear and nonconvex due to physical laws, (b) computational resources required by existing algorithms grow rapidly with the size of the system, and, most importantly, (c) the number of scenarios involving adverse conditions is too large to be expressed (it is higher than the number of atoms in the observable universe for systems with as low as 500 possible attack points). These challenges have limited the scope of previous studies to simple approximate models or conservative methods that ignore the topology-dependent characterization of vulnerabilities [2, 4–24]. Similar hurdles exist in studying the vulnerability of data analytics for other large-scale complex graphs, including ecological and social systems [25, 26], due to the lack of statistical tools for dealing with untrusted data in underlying nonlinear and structured (rather than random) graphical models.

Here, we focus on the U.S. grid, which is the largest machine on earth with more than 450,000 miles of transmission lines (Figure 1). It consists of three large and nearly independent synchronous systems (Eastern, Western, and Texas) that together span the lower 48 states, most of Canada, and some parts of

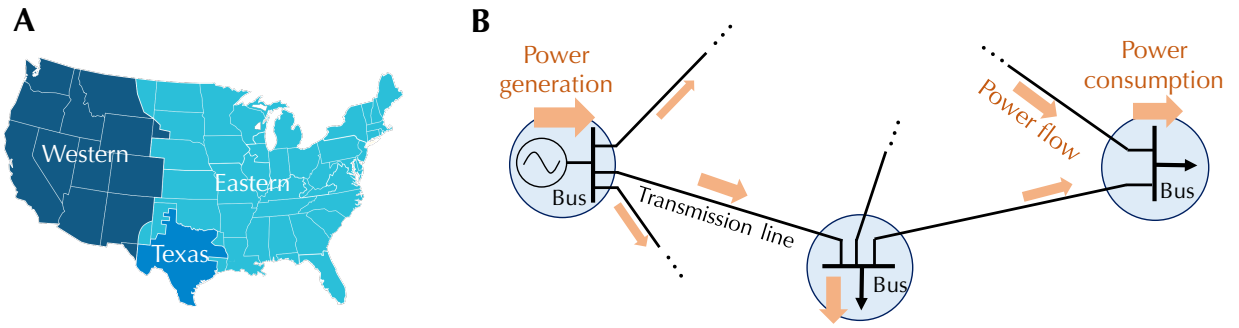

Figure 1: **The U.S. power transmission network. (A)** Map of the Eastern, Western, and Texas interconnections. **(B)** Schematic diagram of a portion of the network. Each blue circle indicates a node (e.g., generator bus or load bus). Nodes are connected by transmission lines. Power is generated, transported, and consumed in different locations (the amount of power is represented by the width of the orange arrow).

Mexico. Due to the confidentiality requirements on critical infrastructure information, we report our findings on somewhat modified grids, which match the size, complexity, and characteristics of actual grids [27].

Power system state estimation

Problem formulation. Power system state estimation aims to find an n-dimensional complex voltage vector $\boldsymbol{v}$ consisting of the voltages at all buses of the network based on a set of measurements, such as voltage magnitude measurements, real and reactive power injections at buses, and power flows over lines. Using the laws of physics, these quantities can be expressed as quadratic functions of $\boldsymbol{v}$ in the form of $\boldsymbol{v}^* \boldsymbol{M}_i \boldsymbol{v}$, where $\boldsymbol{v}^*$ indicates the complex conjugate, and $\boldsymbol{M}_i$ is a known $n \times n$ dimensional Hermitian matrix whose entries depend on the type and location of measurement y_i. As a consequence, each measurement y_i can be written as:

$$y_i = \boldsymbol{v}^* \boldsymbol{M}_i \boldsymbol{v} + \omega_i + b_i \tag{1}$$

for $i \in \{1, ..., m\}$, where ω_i denotes a zero-mean Gaussian random noise that often arises from measurement probes, and b_i denotes bad data that can take arbitrary values. The bad data b_i could originate from cyberattack, communication failure, sensor fault, or deployment of a model $\boldsymbol{M}_i$ that does not match the reality (e.g., a disconnected line is wrongly assumed to be in service by the operator). The goal is to estimate $\boldsymbol{v}$ from $y_1, ..., y_m$, without having any knowledge about ω_i's and b_i's.

Literature review. There are two main challenges for solving the SE problem: nonconvexity (which arises from the quadratic measurement equations and results in the existence of potentially many local minima) and bad data (which can arbitrarily skew the SE solution). Based on how the nonconvexity is addressed, the existing methods can be grouped into three categories: (i) DC approximation, (ii) local search, and (iii) global optimization. Methods in Category (i) approximate each quadratic expression with a linearized model, which transforms the nonconvex SE into a convex optimization [4–6, 10, 28]. However, the approximation error could be arbitrarily large, especially when the unknown voltage deviates from the nominal state around which the linearizion is performed and this becomes signicificantly worse in presence of strategically designed bad data. Category (ii) includes iterative algorithms that solve the nonconvex regression problem formulated according to the quadratic equations. The common choice for the regression objective is a quadratic loss function, resulting in the canonical nonlinear least square (NLS) problem, which can be solved by Newton's method [29, 30] or feasible point pursuit [31]; other choices such as absolute value loss and Huber loss have also been investigated [2, 6, 22]. The challenge, nevertheless, is that local search methods can become trapped at meaningless local minima or saddle points, which are spurious and do not correspond to a useful estimate of the state. A variant of this problem, named matrix completion, has been extensively studied in the machine learning community [32, 33]. Various theoretical conditions on the performance of local search methods to recovering a global optimum of those problems in machine learning have been developed, such as the restricted isometry property [32]; however, those conditions only apply to dense and random matrices $\boldsymbol{M}_i$'s, which is not the case for power systems since $\boldsymbol{M}_i$'s are sparse, structured, and deterministic. The main difference between learning/estimation in power systems and those in machine learning is the existence of inherent structure (captured through the notion of treewidth in graph theory [34]) and lack of randomness. Algorithms in Category (iii) aim to find the global solution of SE. However, the primary disadvantage of these methods, such as particle swarm optimization [35], homotopy continuation methods [36], and semidefinite relaxation [17, 19, 20], is their heavy computational requirement or lack of theoretical guarantees on their ability to find the true state. Moreover, due to the complexity of global optimization techniques, their performance under a worst-case attack scenario is not studied. To

address bad data, efforts in the power system community have been primarily limited to algorithms in Category (i) under the name of bad data detection (BDD) [4–6, 10]. Existing literature on vulnerability and defense of cyberattack, such as the false data injection attack, has also been limited to DC approximation models [7–9, 11, 12, 18], with the exception of a few works on the AC model [16, 37, 38]. However, it has been found that the mismatch caused by the DC approximation of the AC grid renders either the defense or the attack efforts futile [13, 15, 16]. Robust state estimation techniques for nonlinear measurements proposed in the literature include composite optimization [22], iterative mixed ℓ_1 and ℓ_2 convex program [14], semidefinite programming [17, 19, 20], and linear/quadratic programming [23, 24]. Nevertheless, the theoretical conditions in [14, 20] are difficult to verify, and the bound on estimation error in [19] does not apply to bad data rejection. Moreover, none of the existing methods are able to provide a vulnerability measure for each line of the grid that is independent of the unknown attack scenario, because they rely on the locations of the attacked points.

Our contributions. One common drawback of the existing methods is that the theoretical certificates used to reject bad data are provided on a scenario-by-scenario basis, where each scenario corresponds to one specific set of measurements that are corrupted by bad data. Since there are an exponential number of ways to attack the grid data (namely, 2^m ways in the case with m measurements), it is impossible to make a meaningful general assessment of the vulnerability of a grid based on a single scenario. Another important missing factor is that the prior literature aims to find the state of the system correctly under attacks, while this is theoretically impossible when the data for a sub-network of the system is strategically manipulated. In this case, the state for that region becomes unobservable (not recoverable) from the clean data for the rest of the system. To elaborate, let $\boldsymbol{v}$ and $\hat{\boldsymbol{v}}$ be the true and the estimated states, respectively. Let R denote the subnetwork under a cyberattack, and $\boldsymbol{v}_R$ and $\boldsymbol{v}_{\setminus R}$ be the voltages for the attacked region and the remainder of the system, respectively. The existing works (such as [14, 20, 23, 24, 39]) aim to find $\hat{\boldsymbol{v}}$ such that the global metric $\|\boldsymbol{v} - \hat{\boldsymbol{v}}\|$ is minimized; however, this is not possible for real-world attacks since an intelligent attack creates an unobservability issue in the model that relates the clean data to the entire system, which makes the error $\|\boldsymbol{v}_R - \hat{\boldsymbol{v}}_R\|$ always stay significant. In this work, if $\boldsymbol{v}_{\setminus R}$ can be recovered correctly independent of how the attacker has manipulated the data for R, we state that the lines on the boundary of the region R do not allow the estimation error inside the attacked region to be propagated to the rest of the network. **This paper is the first work to develop a mathematical framework for finding the attacked region R together with the state of the system in the uncompromised part of the grid, namely $\boldsymbol{v}_{\setminus R}$, instead of deriving restrictive conditions for finding the true state at all buses.** Our method also provides a vulnerability map for any power grid, as shown in Figure 2 for the U.S. grid. Based on the graphical mutual incoherence condition to be discussed next, we can categorize each line as either robust or vulnerable. On this map, if the connections between the region R and the rest of the grid are all robust lines, then no matter how the measurements in this region, such as the power injections, voltage magnitudes, and power flows are modified, the estimation error is only limited to this region and cannot propagate out of the boundary formed by the robust lines to affect the rest of the grid in terms of $\|\boldsymbol{v}_{\setminus R} - \hat{\boldsymbol{v}}_{\setminus R}\|$. If even one line in the surrounding subnetwork is vulnerable, then it is possible for the estimation error to propagate to the rest of the grid. This vulnerability map is obtained without knowing the attack locations, and therefore it provides a universal measure that applies to an exponential number of possible attack scenarios.

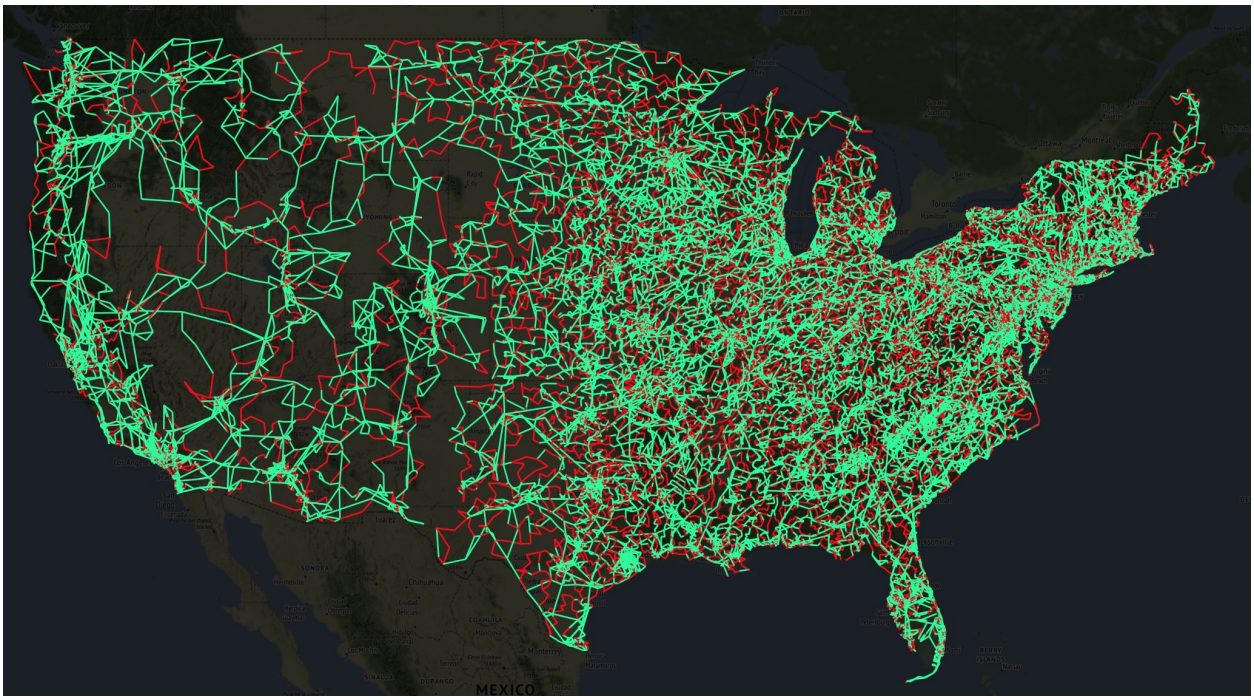

Figure 2: **Vulnerability map of the modified U.S. power grid**, showing the robust (green) and vulnerable (red) lines. This map corresponds to the case in which there are measurements of voltage magnitude and power injections for each bus and there are measurements of real and reactive power flows for each line.

Boundary defense mechanism

Two-step pipeline for robust SE. A power network is modeled as a graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ with the bus set $\mathcal{N}$ and the line set $\mathcal{L}$. In this paper, we focus on a two-step pipeline for robust SE. It is based on a linear representation in a lifted space (subject to a low-rank constraint that will be ignored in the method), which can express common types of measurements, such as the real and reactive power flows and voltage magnitudes, as a linear function of variables [23, 24]. To obtain this representation, we compute the joint sparsity pattern $\boldsymbol{S}$ of the measurement matrices $\boldsymbol{M}_i$ in such a way that the (k,t)-th element $[\boldsymbol{S}]_{k,t}$ is nonzero if and only if there exists an index $i \in \{1, ..., m\}$ with a property that $[\boldsymbol{M}_i]_{k,t}$ is nonzero. We consider a linear basis that is obtained by collecting those variables of $\boldsymbol{v}\boldsymbol{v}^*$ that are at locations corresponding to the nonzero elements in $\boldsymbol{S}$. We denote this set of variables as $\boldsymbol{x}$, which includes the voltage magnitude squares $x_k^{\mathrm{mg}} = |v_k|^2$ for each bus $k \in \mathcal{N}$ as well as $x_\ell^{\mathrm{re}} = \mathbb{R}(v_k v_t^*)$ and $x_\ell^{\mathrm{im}} = \mathbb{I}(v_k v_t^*)$ for each line $\ell = (k,t)$ for which $[\boldsymbol{S}]_{k,t}$ is nonzero, where $\mathbb{R}(\cdot)$ and $\mathbb{I}(\cdot)$ denote the real and imaginary parts of a complex number. Now, one can represent the measurements as:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w} + \boldsymbol{b}. \tag{2}$$

Here, $\boldsymbol{y}$ is the set of m sensor measurements , $\boldsymbol{A}$ is a reconstructed sensing matrix lifted to a high dimension, $\boldsymbol{w}$ is the dense random noise due to measurement errors, and $\boldsymbol{b}$ is the bad data vector with the support $\mathrm{supp}(\boldsymbol{b})$. We assume that $\begin{bmatrix}\boldsymbol{A} & \boldsymbol{I}_{\mathrm{supp}(\boldsymbol{b})}\end{bmatrix}$ has full row rank, where $\boldsymbol{I}_{\mathrm{supp}(\boldsymbol{b})}$ is obtained by selecting those columns of the m-dimensional identity matrix that correspond to the support set $\mathrm{supp}(\boldsymbol{b})$. The elements of $\boldsymbol{x}$ are correlated with each other via the hidden state $\boldsymbol{v}$.

The first step in SE is to solve a convex optimization to minimize the Huber loss of the regression error,

$$\min_{\boldsymbol{x}\in\mathcal{X}} \sum_{i=1}^{m} f_{\text{Huber}}(y_i - [\boldsymbol{A}\boldsymbol{x}]_i; \psi), \tag{3}$$

where $f_{\text{Huber}}(r;\psi) = \begin{cases} \frac{1}{2}r^2 & |r| \le \psi \\ \psi(|r| - \frac{1}{2}\psi) & |r| > \psi \end{cases}$ is the standard Huber loss parametrized by ψ. The variable $\boldsymbol{x}$ can be assumed to be unconstrained, which results in a quadratic programming (QP) problem, as in [23, 24]. However, since the variables correspond to physical quantities, they can be mathematically constrained within a set of second-order cones (SOCs) to improve robustness, which results in a second-order cone programming (SOCP) problem. Based on the solution of Step 1, the next step reconstructs the voltage phasors: the voltage magnitude $|\hat{v}_k|$ is given by $\sqrt{x_k^{\text{mg}}}$, and the phase difference $\hat{\theta}_k - \hat{\theta}_t$ is given by $\arctan\left(\frac{x_\ell^{\text{im}}}{x_\ell^{\text{re}}}\right)$. In an ideal scenario with no dense noise or bad data, one can identify the state uniquely as long as the subgraph associated with the adjacency matrix $\boldsymbol{S}$ forms a spanning tree of $\mathcal{G}$.

Graphical mutual incoherence. We are concerned with the scenario where the data for an entire subregion are compromised, which often arises from natural hazards, equipment faults, and even cyberattacks. We consider the worst-case situation where the attacker can intelligently manipulate every sensor within the region under attack to evade bad data detection. In this situation (illustrated in Figure 3), Newton's method (as the common method used in the industry) is particularly vulnerable because the influence of bad data will propagate throughout the system when solving a nonlinear least-squares problem; on the contrary, the two-step pipeline is completely robust to this situation. To formalize this, we propose a new notion for network defense, referred to as "boundary defense mechanism." For a given attack scenario, there is a natural partition of the network into the attacked region, inner and outer boundaries, and safe region (Figure 4(A)). If the boundary defense is successful, then no matter how erroneous the state estimation within the attacked region is, the estimates at the boundary and in the safe region are unaffected. This framework is fairly general because it incorporates a wide range of adversarial scenarios that are localized, including line outage (where the attack region consists of the two end buses of the line), down substation (where the attack region is a single bus), natural disasters (where the attack region consists of the set of geographically localized buses), and cyberattacks (where the attack region consists of the buses managed by some utility company). Most importantly, we propose a metric called the "graphical mutual incoherence" (gMI) that can be verified for each line in the network to address a large number of possible scenarios on a single map. Note that gMI is defined independently of what part of the network is attacked. For each line $i \to j$, we define i as a node in the attacked region and j as a node in the boundary region (and vice versa for the reverse direction). We use $\mathcal{M}_{\text{bd}\checkmark}^{i\to j}$ and $\mathcal{M}_{\text{bd}\times}^{i\to j}$ to denote the index sets of the defending and defective measurements on the boundary, respectively. Further, $\mathcal{X}_{\text{bd}}^{i\to j}$ indicates the set of variables associated with the boundary, and $\mathcal{L}_{\text{bd}}^{i\to j}$ denotes the set of lines on the boundary (more details can be found in the supplementary material). For the case of unconstrained QP in the first step, gMI for line $i \to j$ is given by the following minimax optimization:

$$\alpha_{i\to j}^{\text{LP}} = \max_{\|\boldsymbol{\xi}\|_\infty \le 1} \quad \min_{\boldsymbol{h}\in\mathcal{H}^{\text{LP}}(\boldsymbol{\xi})} \quad \|\boldsymbol{h}\|_\infty \tag{gMI-QP}$$

where $\mathcal{H}^{\text{LP}}(\boldsymbol{\xi}) = \left\{\boldsymbol{h} \mid \boldsymbol{A}^\top_{\mathcal{M}_{\text{bd}\checkmark}^{i\to j}, \mathcal{X}_{\text{bd}}^{i\to j}} \boldsymbol{h} + \boldsymbol{A}^\top_{\mathcal{M}_{\text{bd}\times}^{i\to j}, \mathcal{X}_{\text{bd}}^{i\to j}} \boldsymbol{\xi} = \boldsymbol{0}\right\}$ is the set of admissible $\boldsymbol{h}$ for a given vector $\boldsymbol{\xi}$ in the unit hypercube. Here, $\boldsymbol{\xi}$ can be regarded as the subgradient of the bad data vector, and $\boldsymbol{h}$ is an auxiliary

variable associated with good data. We use the subscript notation $\boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}\checkmark}^{i\to j},\mathcal{X}_{\mathrm{bd}}^{i\to j}}$ to indicate the submatrix of $\boldsymbol{A}$ whose rows are indexed by $\mathcal{M}_{\mathrm{bd}\checkmark}^{i\to j}$ and columns are indexed by $\mathcal{X}_{\mathrm{bd}}^{i\to j}$. Figure 4(B) illustrates the nodes and lines relevant to the evaluation of four lines for a given attack scenario. The case of incorporating SOC in the first step is treated similarly:

$$\alpha_{i\to j}^{\mathrm{SOCP}}(\boldsymbol{x}) = \max_{\|\boldsymbol{\xi}\|_\infty \le 1} \quad \min_{\boldsymbol{h}\in\mathcal{H}^{\mathrm{SOCP}}(\boldsymbol{\xi},\boldsymbol{x})} \quad \|\boldsymbol{h}\|_\infty \tag{gMI-SOC}$$

where $\mathcal{H}^{\mathrm{SOCP}}(\boldsymbol{\xi},\boldsymbol{x})$ is the set of admissible $\boldsymbol{h}$. A vector belongs to this set if there exists a series of nonnegative coefficients $\{\omega_\ell\}_{\ell\in\mathcal{L}_{\mathrm{bd}}^{i\to j}}$ such that $\boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}\checkmark}^{i\to j},\mathcal{X}_{\mathrm{bd}}^{i\to j}}^{\top}\boldsymbol{h} + \boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}\times}^{i\to j},\mathcal{X}_{\mathrm{bd}}^{i\to j}}^{\top}\boldsymbol{\xi} + \sum_{\ell\in\mathcal{L}_{\mathrm{bd}}^{i\to j}}\omega_\ell \boldsymbol{T}_\ell \boldsymbol{x} = \boldsymbol{0}$, where $\boldsymbol{T}_\ell$ is a coefficient matrix defined for every line $\ell \in \mathcal{L}_{\mathrm{bd}}^{i\to j}$ (see the supplementary material for more details).

Firstly, it can be seen that (gMI-SOC) depends on the true state $\boldsymbol{x}$. However, this dependence is not an issue because it can be shown that for every $\boldsymbol{x}$ that corresponds to a complex voltage state of the system, it holds that

$$\alpha_{i\to j}^{\mathrm{SOCP}}(\boldsymbol{x}) \le \alpha_{i\to j}^{\mathrm{LP}}$$

In other words, the incorporation of second-order cone constraints *always* improves robustness. By definition, for a network with n_ℓ lines, we only need to evaluate $2 \times n_\ell$ gMI indices (two for each line), which is independent of the attack scenarios. For computational efficiency, even though (gMI-QP) and (gMI-SOC) are defined as min-max problems, the outer maximization problems can be solved efficiently by enumeration over the power set of the measurements located at the boundary (including power flows over the line from bus j (boundary node) to bus i (attacked node) and power injections at bus j), which has at most 16 points for each line. We have also developed an algorithm by reformulating the min-max optimization as a linear complementarity problem or a mixed-integer problem, as detailed in the supplementary material.

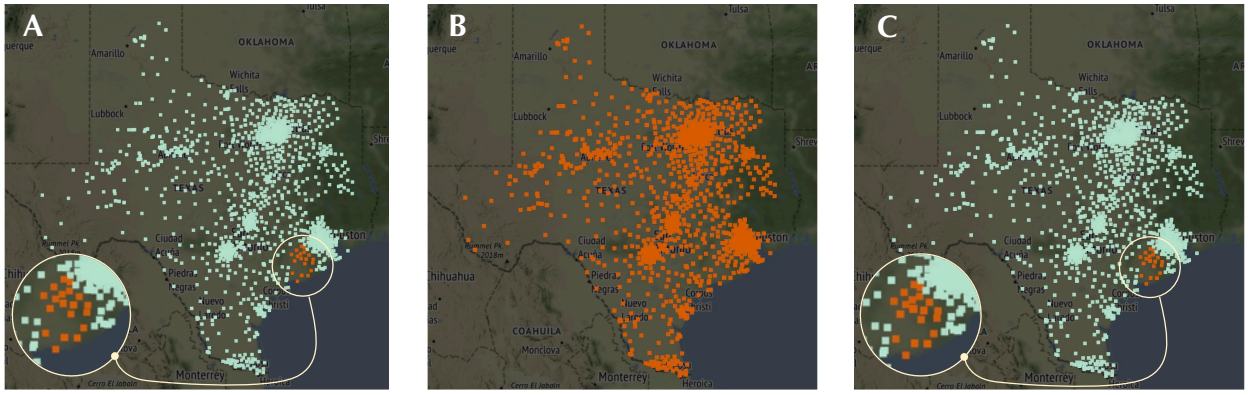

Figure 3: **Evaluation of the boundary defense mechanism. (A)** The grid is under "zonal attack," where the measurements within a zone are corrupted (shown in red). State estimation based on **(B)** Newton's method for nonlinear least squares, and **(C)** the proposed method with SOC constraints, where in both cases, buses with an estimation error greater than 0.002 are marked in red. The errors propagate throughout the grid in **(B)** but are contained within the zonal boundary in **(C)**.

Theorem 1 (Boundary defense mechanism)**.** *Consider a partition of the network into the attacked, boundary, and safe regions, where the bad data are contained within the attacked region. Assume that the following two conditions are satisfied:*

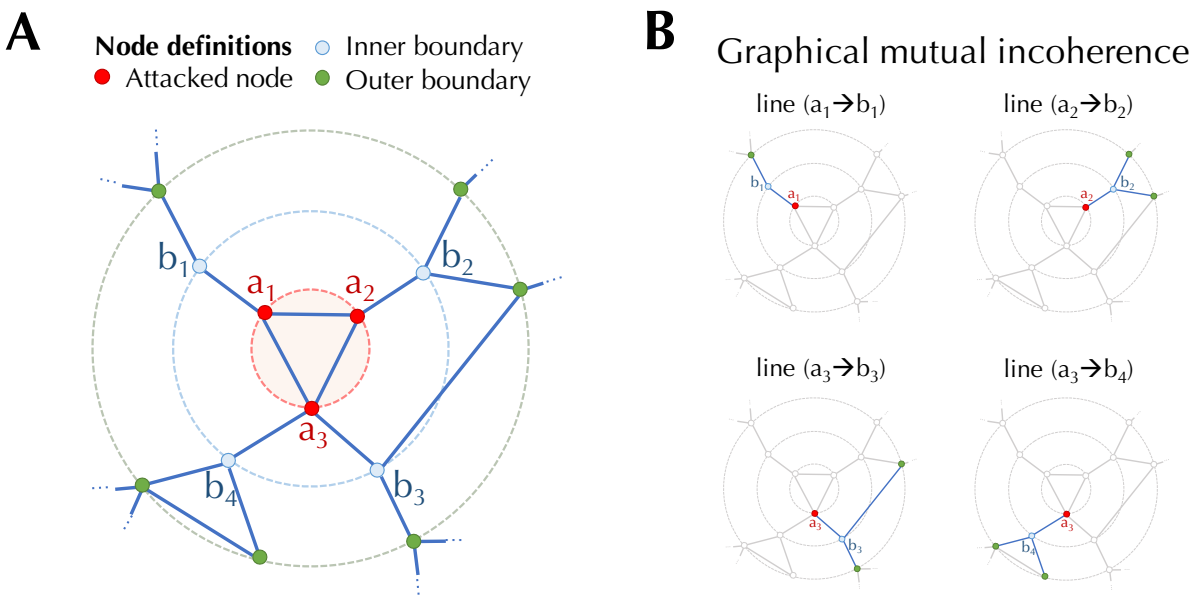

Figure 4: **Illustration of the boundary defense mechanism. (A)** Schematic diagram showing the attacked nodes as well as inner and outer boundary nodes. **(B)** Graphical mutual incoherence evaluation. Only nodes and lines considered in the evaluation are highlighted for each line evaluation, with each line direction considered to be from the attacked node to the inner boundary node.

- *Full column rankness condition:* $\boldsymbol{A}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bd},\mathcal{X}_{\rm sf}\cup\mathcal{X}_{\rm bd}}$ *and* $\boldsymbol{Q}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}} = \begin{bmatrix}\boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}} & \boldsymbol{I}_{\mathcal{M}_{\rm bi}}^{(|\mathcal{M}_{\rm bd}|)\top}\end{bmatrix}$ *have full column rank, where* $\mathcal{M}_{\rm sf}$*,* $\mathcal{M}_{\rm bi}$ *and* $\mathcal{M}_{\rm bo}$ *are the sets of measurements in the safe region, inner boundary and outer boundary, respectively,* $\mathcal{M}_{\rm bd} = \mathcal{M}_{\rm bi}\cup\mathcal{M}_{\rm bo}$ *is the boundary measurements,* $\mathcal{X}_{\rm sf}$ *and* $\mathcal{X}_{\rm bd}$ *are variables in the safe region and boundary,* $|\mathcal{M}_{\rm bd}|$ *is the number of measurements in the boundary, and* $\boldsymbol{I}_{\mathcal{M}_{\rm bi}}^{(|\mathcal{M}_{\rm bd}|)}$ *is the submatrix that consists of the* $\mathcal{M}_{\rm bi}$ *columns of the* $|\mathcal{M}_{\rm bd}|$*-dimensional identity matrix.*
- *Graphical mutual incoherence: For every line* $\{i,j\}$ *that bridge the attacked region and the inner boundary, where* i *and* j *are in the attacked region and inner boundary, respectively, it holds that* $\alpha_{i\to j} < 1-\gamma$ *for some* $\gamma > 0$*.*

Then, the solution obtained from the two-step pipeline has two properties: ***(i)*** *every measurement flagged by the algorithm correctly belongs to bad data, so there are no false positives in Step 1; and* ***(ii)*** *after removing the subgraph of the attacked region from the main graph, direct recovery in Step 2 recovers the underlying state of the system for the region that has not been attacked.*

The above result holds true in the general case with dense noise for every measurement, except that only those bad data that exceed a threshold can be detected with guarantees, and therefore those data below the threshold are treated as dense noise. We relegate the details to the supplementary material. The above theorem provides a modular approach for constructing theoretical guarantees for defense against bad data. Various security indices have been proposed in the literature for DC [8, 9, 12] and AC SE [16, 37, 38]. However, they are based on a single attack plan generated by an optimization problem and cannot be meaningfully applied to an exponential number of other attacks. On the contrary, with our method, the evaluation

of the graphical mutual incoherence for each line is independent of the attack scenarios. As a consequence, a boundary defense is established as long as there exists a subgraph enclosing the attacked region that satisfies the conditions in Theorem 1, such as in the scenario illustrated in Figure 3. It is also worth mentioning that we do not distinguish the causes of bad data, such as equipment failures, misrepresenting the physical model due to time-varying changes in the network, or cyberattacks. The calculation of gMI also does not depend on the sparsity of the attack vector.

Geographic mapping of vulnerabilities

Vulnerability map. Based on the mathematical tools developed in the previous section, we assess the robustness of the synthetic U.S. grid. First, we visualize the gMI on the map for both (gMI-QP) and (gMI-SOC) in Figure 5. Due to its dependence on the underlying state, (gMI-SOC) is shown for a profile described by the dataset, which represents a snapshot of the operating status. This snapshot only serves as an illustration of how much improvement in robustness is achieved by incorporating SOCs into a typical operating condition. A line is considered "robust" if the gMIs in both directions are less than 1; otherwise, it is "vulnerable" (a V-line). The plot shows a geographic distribution of robust/vulnerable lines for the Eastern U.S. grid. It can be seen that the density of vulnerable lines is relatively high for populated areas, such as Boston and New York, where we also observe a high density of robust lines. On average, 59% lines are robust across the states, which are then split further into independent synchronous regions, as shown in Table 1. In addition, the map validates that (gMI-SOC) always improves (gMI-QP), which implies that the incorporation of SOCs can help rectify state estimations and detect bad data.

The vulnerability map can be used in various ways. For instance, it can be used to investigate whether topological errors for a line or a substation can be contained locally, corresponding to the case in which there

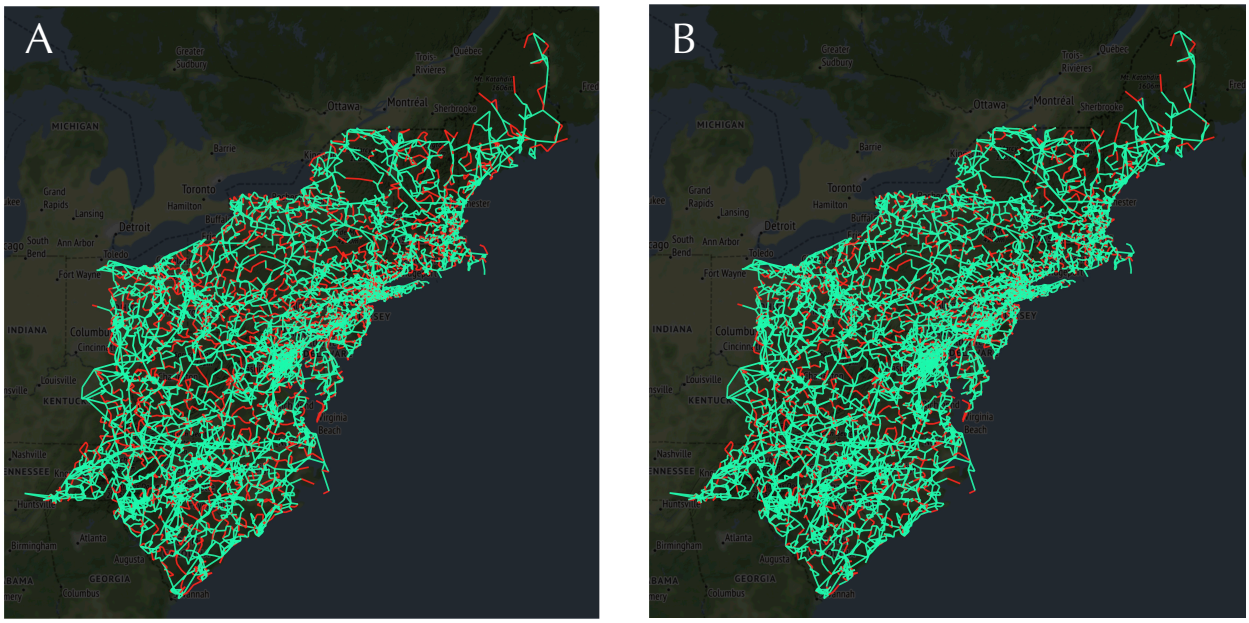

Figure 5: **Comparison of vulnerability maps under different optimization strategies.** Vulnerability maps when using the proposed **(A)** QP and **(B)** SOCP are shown, where robust lines are marked in green, and vulnerable lines are marked in red.

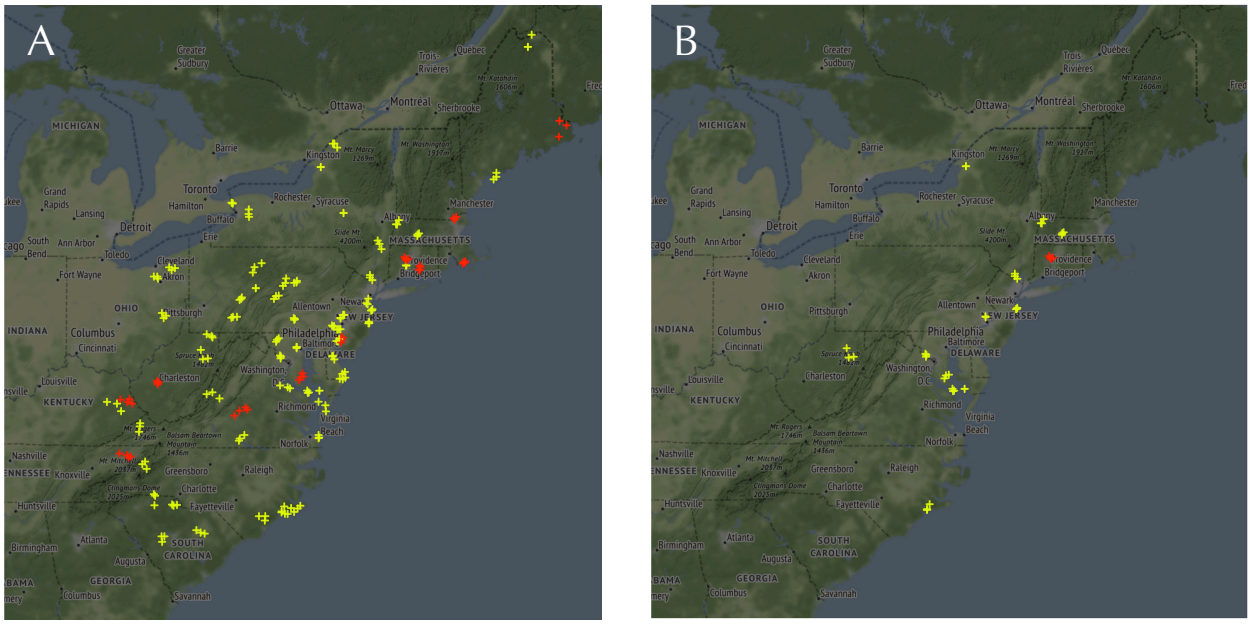

Figure 6: **Comparison of bus critical index maps under different optimization strategies.** Since the bus critical indices are no larger than 3 within the map, we only show the locations with values 2 (yellow) and 3 (red) for the proposed **(A)** QP and **(B)** SOCP state estimation strategies.

is a model mismatch for a transmission line or substation, such that the associated measurements are largely biased. While this is a challenging problem, it could be addressed systematically using the vulnerability map. Specifically, if the erroneous line/substation is surrounded by robust lines, then it is guaranteed that the error will be contained locally via the boundary defense mechanism. Otherwise, there is a possibility that the error will "escape" through a vulnerable line, which is referred to as a "critical line (C-line)" or a "critical bus (C-bus)," to affect the outside region. In particular, for topological errors such as line misspecification, it can be regarded as a pair of gross injection errors at the two ends of the line; hence, we can identify it as long as the line is not a C-line. Summary statistics are shown in Table 1.

Criticality index for substations under cyberattack. Furthermore, we can extend the case study by defining a criticality index (CI) for each substation. The CI gauges how many nodes in a substation's neighborhood will be affected if it is down. The higher the value, the more crucial the situation is when the substation is compromised. This situation is analogous to the cascading failures of generators, but the difference is clear—our focus is on the robustness of the data analytics rather than the physical dynamics. For each substation, the CI can be calculated as the size of the connected component rooted at the node, where an edge between two nodes is present if and only if the physical line that connects them is vulnerable. We visualize the distribution of the CIs on the map shown in Figure 6, and it can be seen that they are concentrated in populated areas.

Relating vulnerability to network and optimization properties

Measurement types and locations. To investigate factors that affect gMI, we shift our focus to the underlying network and optimization properties. So far, our study has been conducted with respect to a specific

	Basic properties		Properties of QP				Properties of SOCP			
	Buses	Lines	V-lines	C-lines	C-bus	Bus CI	V-lines	C-lines	C-bus	Bus CI
Texas	2,000	3,206	.3762	.4251	.4775	.20	.2979	.3674	.4225	.06
Western	10,000	12,706	.4715	.5231	.5313	.15	.3979	.4636	.4860	.06
Eastern	70,000	88,207	.4932	.5415	.5327	.14	.4104	.4780	.4810	.05

Table 1: **Summary statistics of network properties and vulnerability characteristics.** We show the percentage of V-lines and C-lines among all network lines, and the percentage of C-buses among all network buses for QP and SOCP. We also show the average bus critical index, which measures the influence of a single-bus attack on the rest of the network.

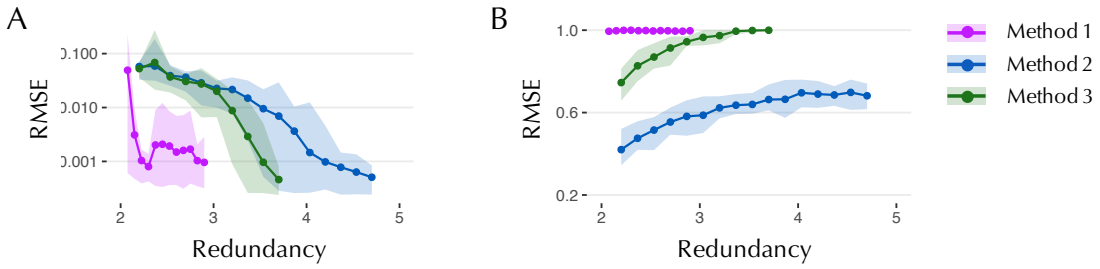

Figure 7: **Comparison of different measurement profiles and redundancy.** We consider three different methods for sensor augmentation, as detailed in the main text. The redundancy value is calculated as the number of sensors divided by $2 \times n_b$(number of buses) $- 1$, which represents the degrees of freedom in the traditional power flow problem. Each point for the **(A)** RMSE and **(B)** F1 score is obtained by averaging over 100 independent simulations. The average value is shown by the solid line, and the 5% and 95% quantiles are shown by the shaded region.

measurement profile which corresponds to the set of full nodal and branch measurements. Important question are: *How do the number and locations of measurement sensors affect line vulnerability?* In particular, does decreasing the number of sensors make the network significantly more vulnerable? What type of sensor measurements can bolster boundary defenses?

For this purpose, we examine three methods used for "measurement augmentation." The first method (Method 1) starts from a spanning tree of the network and adds a set of lines to the tree incrementally to obtain a subgraph that will be used for taking measurements. In this method, each bus is equipped with only voltage magnitude measurements, and each line has three out of four branch flow measurements. The second method (Method 2) starts with the full network, where each node has voltage magnitude measurements, and each line has one real and one reactive power measurement, and it grows the set of sensors by randomly adding branch measurements. The third method (Method 3) differs from Method 2 only in that it grows the set of sensors by randomly adding branch measurements as well as nodal power injections. To evaluate these three methods, we devise a "scattered attack" strategy, where we randomly select 25 lines from the 2000-bus Texas interconnection and corrupt all of its branch measurements, which amounts to 100 bad pieces of data. We then employ our proposed method to first detect bad data, and then rerun SE on the sanitized measurement set. The observation is that, in general, both the root mean squared error (RMSE) and the F1 score for bad data detection are enhanced as more sensors are added to the network, as shown in

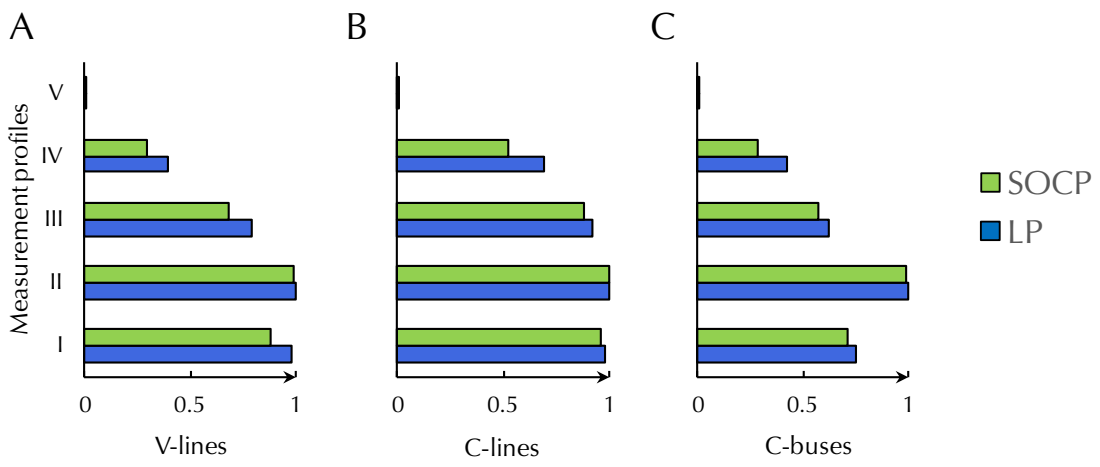

Figure 8: **Characterization of vulnerability based on measurement profiles.** The five measurement profiles used are: full nodal measurements and two/three/four branch flows per line (I/III/IV); real and reactive power injections per bus and three branch flows per line (II); and voltage magnitude per bus and three branch flows per line (V). For each state estimation method (QP or SOCP), we show the percentage of **(A)** V-lines, **(B)** C-lines, and **(C)** C-buses within the Texas interconnection.

Figure 7. The F1 score is given by $\frac{2\text{precision}*\text{recall}}{\text{precision}+\text{recall}}$, where precision is the rate of true positives (i.e., correctly identified bad data) among all data that are claimed to be bad, whereas recall is given by the percentage of true positives identified as bad data) among all ground truth bad data. Specifically, an F1 score close to 1 indicates that the algorithm detects all bad data (high recall rate) and does not falsely blame the good data (high precision rate).

There is also a major discrepancy among the different methods at the same level of measurement redundancy. For instance, Method 1 significantly outperforms the other two methods at a low redundancy rate, whereas Method 2 steadily outmatches Method 3 with more sensors. To explain this phenomenon, we need to examine the types of available measurements. Thus, we select five typical measurement profiles as snapshots of Figure 7 and calculate the percentage of V-lines and C-lines, and the average CI in each case (Figure 8). It turns out that the inclusion of voltage magnitude or branch flow measurements can enhance the robustness, whereas the addition of nodal power injections is a major factor in weakening the defense. For example, with only voltage magnitude and branch flow measurements, the network is almost "everywhere defendable," namely, the locations of scattered attack can be detected accurately with high probability. On the contrary, with the inclusion of nodal injections, even with a high rate of branch flow measurements, the network is still vulnerable. Intuitively, this situation occurs because nodal power injections are highly coupled measurements which depend on state variables for all lines connected to the node. When one or a few of the branches are under attack, this scenario can lead to miscalculations for all incident lines. In contrast, voltage magnitudes and branch flows are more localized in nature, and, when corrupted, they have fewer effects on adjacent buses/lines.

Topological properties. In addition to the measurement set, network vulnerability also depends on topological properties. In particular, our findings show that the connectivity degree for each node is positively correlated with line vulnerability (Figure 9(A)). A boundary defense node is increasingly likely to defend against attacks as the degree increases. However, this trend is less obvious when the node is under attack,

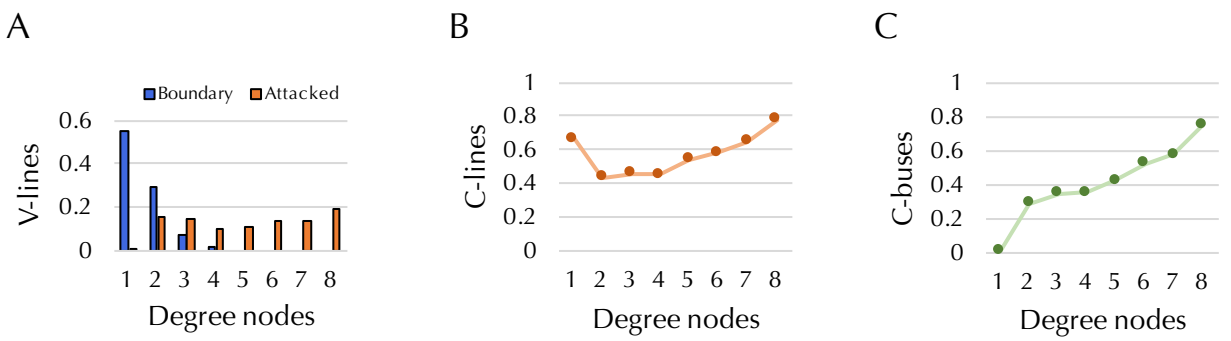

Figure 9: **Characterization of vulnerability through nodal degrees. (A)** Percentage of V-lines when the nodes are at the boundary or in the attacked region. In this case, we distinguish the two directions of a line. Percentage of **(B)** C-lines and **(C)** C-buses averaged over nodes with the same degree. Since the distribution of nodal degrees is light-tailed, we group nodes of degree eight or higher in the same bin.

since high-degree nodes have more measurements from the region not under attack to leverage in order to rectify the corrupted lines. On the other hand, it is more likely that a line will be critical if it is connected to a high-degree bus, as is shown in Figure 9(B). This criticality can be explained via the definition of a critical line, and as long as at least one of the remaining lines incident to that bus is vulnerable, the error will propagate out through that vulnerable line. Similarly, a high-degree node is more likely to be a critical bus. In addition to the degree of connections, which is a local property, we have observed an interesting relation to the tree decomposition of the network, which provides a generalization of the method under discussion. However, due to the technicality of the explanation, we leave this observation to the supplementary materials.

Optimization property. As for the optimization property, our theoretical analysis indicates that the incorporation of SOCs always improves line robustness (Proposition 2S), which can be verified visually in Figure 5 and observed in Figure 8 for different measurement profiles.

Conclusion

Our vulnerability analysis of power system state estimation is distinguished from previous works by its scalability but also by the strong formal guarantees of a boundary defense against cyberattacks and a localized vulnerability assessment that accounts for network and optimization properties. This study provides a set of notions and tools—the development of graphical mutual incoherence, the boundary defense mechanism, and the analysis of topological and optimization relations to vulnerability—that are applicable to a wide range of graph-structured data.

Our analysis is based on the assumption that the amount of data is not too low—an assumption that is far from being restrictive, as it is shown that with the right set of measurements, one can identify the true state of the system with only one more sensor per bus, on average, compared to the classical setting of power flow that is known to have multiple spurious local minima. More importantly, the emerging scenario of "abundant but untrusted data" considered in this study is more practically realistic and algorithmically challenging than the traditional scenario of "redundant and reliable data." Based on a robust two-step algorithm, we developed a boundary defense mechanism to defend against cyberattacks. This defense is drastically different from

those proposed in the existing literature, since it addresses a vast number of attack and defense scenarios via a single vulnerability map. By using this tool, this is the first work that performs a system-wide vulnerability assessment of power networks at the size of the entire U.S. grid.

Based on the proposed mathematical framework, our analysis revealed several key factors that could affect the robustness of a network. A highly connected node is able to defend against attacks if it occurs to lie on the boundary, but it is also more prone to attacks resulting in higher collateral damage. For a given topological structure, the inclusion of nodal power injection data can weaken the defense; by contrast, the inclusion of voltage magnitude or branch power flow measurements can enhance the robustness against bad data, giving rise to a higher bad data detection accuracy. From an algorithmic perspective, the incorporation of second-order cone constraints is theoretically shown to be beneficial for network robustness and validated through extensive experiments. Our analysis offers a scientific foundation for vulnerability-based resource allocation, which, in the case of a power grid, would be based on prioritizing the upgrade of sensing infrastructure for critical locations.

Method summary

The power grid is modeled as a network of buses connected by transmission lines, where each bus is associated with a complex voltage phasor as its state. Given the topology and measurement profile, some linear basis variables can be constructed for each bus and branch adaptively—if there are no branch measurements and nodal power injections on the connected buses, then the corresponding branch variables can be ignored. Doing so ensures the sparsity of the basis. From the measurements, we first estimate the linear basis using quadratic or second-order cone programming. Bad data detection is performed by thresholding the estimated bad data vector. Then we rerun the estimation on the sanitized dataset, and the results are fed into the second step in the pipeline to produce a state estimation.

We consider two types of attacks. The first attack is a "scattered attack" (Figure 7), where a random subset of lines are chosen whose measurements are all corrupted. In this case, the bad data are scattered throughout the network, and the goal is to recover the overall system state correctly. The second attack is a "zonal attack" (Figure 3), where all measurements within a zone—usually governed by a single utility—are corrupted. In this case, the goal is to identify the boundary of the attack and recover the state outside the attacked zone correctly. In the case of a stealth attack, there is a problem of symmetry, namely, without additional information, it is impossible to decide which zone is under attack since the only inconsistency is observed at the boundary. To avoid this case, we arbitrarily break the symmetry by introducing some sensors within the attacked zone that are more secure than others in such a manner that their values cannot be modified. We can also perform posterior inference based on our prior knowledge of which zones are more likely to be secure.

The vulnerability analysis is based on the partition of measurements and variables into attacked and boundary categories (Figure 4). The graphical mutual incoherence is defined by a min-max problem which is NP-hard in general. However, this set-up does not pose a computational challenge since the gMI for each line can be calculated through an efficient enumeration strategy that scales exponentially according to the number of bad measurements that is limited for a line. The gMIs for the entire U.S. grid (more than 100,000 lines) can be obtained in less than a day on a personal computer. For a large-scale instance, we propose two reformulations of the problem, namely, a linear complementarity problem and mixed-integer programming, which can be employed to solve the problem efficiently. The critical index for buses (Figure 6) is obtained by counting the size of the subgraph rooted at the substation and linked by a directional edge that is vulnerable. A critical line is identified when any one of the adjacent lines pointing outwards is vulnerable.

The formal result of the boundary defense mechanism is established through a series of propositions and lemmas. The key steps include: (1) a "glueable property," which shows that a local graphical mutual incoherence property implies a global property (Lemmas 9S and 15S in the supplementary material), (2) a result that establishes that a boundary defense can stop error propagation (Lemmas 5S and 12S), and (3) a statistical analysis of the first-step algorithm based on concentration bounds and a primal-dual witness argument (Theorems 10S, 11S, 17S, and 18S). Further details on the linear representation, two-step pipeline algorithm, theoretical analysis, and experimental setup are given in the supplementary materials.

Author contributions statement

M.J. and J.L. developed the idea. M.J. developed the theoretical formalism. M.J., J.L., R.B. and S.S. designed the experiments and interpreted the results. M.J. performed the experiments, analyzed the data and prepared the manuscript. S.S., J.L. and R.B. revised the manuscript.

References

[1] Engineering National Academies of Sciences, Medicine, et al. *Enhancing the resilience of the Nation's electricity system*. National Academies Press, 2017.

[2] Ali Abur and Antonio Gomez Exposito. *Power system state estimation: theory and implementation*. CRC press, 2004.

[3] U.S.-Canada Power System Outage Task Force. Final report on the August 14, 2003 blackout in the United States and Canada: Causes and recommendations. 2004.

[4] H. M. Merrill and F. C. Schweppe. Bad data suppression in power system static state estimation. *IEEE Transactions on Power Apparatus and Systems*, PAS-90(6):2718–2725, 1971.

[5] Willy W Kotiuga and M Vidyasagar. Bad data rejection properties of weighted least absolute value techniques applied to static state estimation. *IEEE Transactions on Power Apparatus and Systems*, (4):844–853, 1982.

[6] A Monticelli. Electric power system state estimation. *Proceedings of the IEEE*, 88(2):262–282, 2000.

[7] Oliver Kosut, Liyan Jia, Robert J Thomas, and Lang Tong. Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures. In *IEEE International Conference on Smart Grid Communications*, pages 220–225, 2010.

[8] Gyorgy Dan and Henrik Sandberg. Stealth attacks and protection schemes for state estimators in power systems. In *IEEE International Conference on Smart Grid Communications*, pages 214–219, 2010.

[9] Henrik Sandberg, André Teixeira, and Karl H Johansson. On security indices for state estimators in power networks. In *First Workshop on Secure Control Systems*, 2010.

[10] Yih-Fang Huang, Stefan Werner, Jing Huang, Neelabh Kashyap, and Vijay Gupta. State estimation in electric power grids: Meeting new challenges presented by the requirements of the future grid. *IEEE Signal Processing Magazine*, 29(5):33–43, 2012.

[11] Yao Liu, Peng Ning, and Michael K Reiter. False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security*, 14(1):13, 2011.

[12] André Teixeira, György Dán, Henrik Sandberg, and Karl H Johansson. A cyber security study of a scada energy management system: Stealthy deception attacks on the state estimator. *IFAC Proceedings Volumes*, 44(1):11271–11277, 2011.

[13] Wenye Wang and Zhuo Lu. Cyber security in the smart grid: Survey and challenges. *Computer networks*, 57(5):1344–1371, 2013.

[14] Weiyu Xu, Meng Wang, Jian-Feng Cai, and Ao Tang. Sparse error correction from nonlinear measurements with applications in bad data detection for power networks. *IEEE Transactions on Signal Processing*, 61(24):6175–6187, 2013.

[15] Md Ashfaqur Rahman and Hamed Mohsenian-Rad. False data injection attacks against nonlinear state estimation in smart power grids. In *IEEE Power & Energy Society General Meeting*, pages 1–5, 2013.

[16] Jingwen Liang, Lalitha Sankar, and Oliver Kosut. Vulnerability analysis and consequences of false data injection attack on power system state estimation. *IEEE Transactions on Power Systems*, 31(5):3864–3872, 2015.

[17] Yang Weng, Marija D Ilić, Qiao Li, and Rohit Negi. Convexification of bad data and topology error detection and identification problems in ac electric power systems. *IET Generation, Transmission & Distribution*, 9(16):2760–2767, 2015.

[18] Gaoqi Liang, Junhua Zhao, Fengji Luo, Steven R Weller, and Zhao Yang Dong. A review of false data injection attacks against modern power systems. *IEEE Transactions on Smart Grid*, 8(4):1630–1638, 2016.

[19] Yu Zhang, Ramtin Madani, and Javad Lavaei. Conic relaxations for power system state estimation with line measurements. *IEEE Transactions on Control of Network Systems*, 5(3):1193–1205, 2017.

[20] R. Madani, J. Lavaei, and R. Baldick. Convexification of power flow equations in the presence of noisy measurements. *IEEE Transactions on Automatic Control*, 64(8):3101–3116, 2019.

[21] Daniel K Molzahn, Ian A Hiskens, et al. A survey of relaxations and approximations of the power flow equations. *Foundations and Trends® in Electric Energy Systems*, 4(1-2):1–221, 2019.

[22] G. Wang, G. B. Giannakis, and J. Chen. Robust and scalable power system state estimation via composite optimization. *IEEE Transactions on Smart Grid*, 10(6):6137–6147, 2019.

[23] Ming Jin, Igor Molybog, Reza Mohammadi-Ghazi, and Javad Lavaei. Towards robust and scalable power system state estimation. In *IEEE Conference on Decision and Control*, 2019.

[24] Ming Jin, Igor Molybog, Reza Mohammadi-Ghazi, and Javad Lavaei. Scalable and robust state estimation from abundant but untrusted data. *IEEE Transactions on Smart Grid*, 2019.

[25] Alessandro Vespignani. Complex networks: The fragility of interdependency. *Nature*, 464(7291):984, 2010.

[26] Alexander A Ganin, Emanuele Massaro, Alexander Gutfraind, Nicolas Steen, Jeffrey M Keisler, Alexander Kott, Rami Mangoubi, and Igor Linkov. Operational resilience: concepts, design and analysis. *Scientific reports*, 6:19540, 2016.

[27] Adam Birchfield, Ti Xu, Kathleen Gegner, Komal Shetye, and Thomas Overbye. Grid structural characteristics as validation criteria for synthetic networks. *IEEE Transactions on Power Systems*, 32(4):3258–3265, 2017.

[28] R Baldick, KA Clements, Z Pinjo-Dzigal, and PW Davis. Implementing nonquadratic objective functions for state estimation and bad data rejection. *IEEE Transactions on Power Systems*, 12(1):376–382, 1997.

[29] William F Tinney and Clifford E Hart. Power flow solution by newton's method. *IEEE Transactions on Power Apparatus and Systems*, (11):1449–1460, 1967.

[30] Fred C Schweppe and J Wildes. Power system static-state estimation, part i: Exact model. *IEEE Transactions on Power Apparatus and Systems*, (1):120–125, 1970.

[31] Gang Wang, Ahmed S Zamzam, Georgios B Giannakis, and Nicholas D Sidiropoulos. Power system state estimation via feasible point pursuit: Algorithms and cramér-rao bound. *IEEE Transactions on Signal Processing*, 66(6):1649–1658, 2018.

[32] Yuejie Chi, Yue M Lu, and Yuxin Chen. Nonconvex optimization meets low-rank matrix factorization: An overview. *arXiv preprint arXiv:1809.09573*, 2018.

[33] Joshua Comden, Andrey Bernstein, and Zhenhua Liu. Sample complexity of power system state estimation using matrix completion. *arXiv preprint arXiv:1905.01789*, 2019.

[34] Ramtin Madani, Somayeh Sojoudi, Ghazal Fazelnia, and Javad Lavaei. Finding low-rank solutions of sparse linear matrix inequalities using convex optimization. *SIAM Journal on Optimization*, 27(2):725–758, 2017.

[35] Shigenori Naka, Takamu Genji, Toshiki Yura, and Yoshikazu Fukuyama. A hybrid particle swarm optimization for distribution state estimation. *IEEE Transactions on Power Systems*, 18(1):60–68, 2003.

[36] Weimin Ma and James S Thorp. An efficient algorithm to locate all the load flow solutions. *IEEE Transactions on Power Systems*, 8(3):1077–1083, 1993.

[37] Gabriela Hug and Joseph Andrew Giampapa. Vulnerability assessment of ac state estimation with respect to false data injection cyber-attacks. *IEEE Transactions on Smart Grid*, 3(3):1362–1370, 2012.

[38] M. Jin, J. Lavaei, and K. H. Johansson. Power grid AC-based state estimation: Vulnerability analysis against cyber attacks. *IEEE Transactions on Automatic Control*, 64(5):1784–1799, 2019.

[39] Yu Zhang, Ramtin Madani, and Javad Lavaei. Conic relaxations for power system state estimation with line measurements. *IEEE Transactions on Control of Network Systems*, 5(3):1193–1205, 2018.