

Improvement of SCADA-based Preventive Control Under Budget Constraints

Vajiheh Farhadi, Sai Gopal Vennelaganti, Ting He, Nilanjan Ray Chaudhuri, Thomas LaPorta

Abstract—Catastrophic disasters in real-world systems, such as large-scale blackouts in power grids, are usually triggered by minor incidents, which culminate in a complex cascading failure in an interdependent system. Because the loss of a power transmission line disrupts the control information piggybacked on the line, failures in the power network may consequently disrupt monitoring and control of the system. Hence, reliable functioning of the communication network in support of monitoring and control is vital to ensure that the re-dispatch-based preventive control effectively restricts cascade propagation. In this paper, we address this issue by proposing a novel scheme in designing the communication network comprised of both power line carrier communication (PLCC) links and non-PLCC (e.g., microwave) links in preparation of possible failures under a budget constraint on the communication link deployment cost. First, we characterize the fundamental hardness of our problem. Next, we develop a solvable Mixed-integer linear programming (MILP)-based algorithm, which attains a constant-factor approximation under certain conditions. Finally, we show via simulations on the IEEE 118-bus system that the proposed algorithm achieves superior performance in terms of enabling more accurate topology estimation and more served demand in the face of cascades.

Index Terms—DC-QSS, cascading failure, blackouts, preventive control, SCADA, PLCC, non-PLCC

I. INTRODUCTION

Cascading failure is the process in an interconnected system in which damage or loss in one part of the system triggers damage or loss throughout the system. For example, in power grids, failure of one element may lead to overloading of nearby elements that results in tripping of those elements through protective action, which leads to further overloading of other components. This process may continue until a substantial number of elements are overloaded or the entire system collapses. If two interconnected systems interact with each other, cascading failures become even more complex.

The electrical energy grid is a cyber-physical system that is comprised of two interconnected systems. The physical system includes generators, transformers, loads, and transmission lines, whereas the cyber system handles the functions of Supervisory Control and Data Acquisition (SCADA). When a failure propagates, it affects the physical and the cyber systems in a complex and coupled manner. Understanding this phenomenon requires understanding both systems, including their major reasons for failures and the role of the SCADA system.

The efficient and secure operation of power systems relies heavily on the associated SCADA system [1], which is one of the most important communication systems supporting the power industry [2]. The SCADA system contains a Control Center (CC), which collects measurements from sensors and issues preventive control commands to generators and loads. The commands can be transmitted through dedicated wires (e.g., fiber optic or coaxial cables), wireless links (e.g., satellite or microwave), or power line carrier communication (PLCC). A system can also use a combination of these technologies.

PLCC (or PLC) [3]–[5] refers to a communication technology between substations using power lines [6]–[9]. PLCC was initially used for one-way communication to monitor power grids. Today PLCC is used for ripple control [10] and bi-directional communication [11] in the power grid. PLCC is often combined with other non-PLCC technologies, such as GSM or fiber [12], to support communications within the SCADA system. Examples of such include the Siemens PowerLink PLCC system [13] and General Electric T&D Power Utilities [14]. The advantage of PLCC is that it is less expensive than dedicated communication links (including wireless links) [15], [16], which makes it a vital component of the SCADA system [9].

One major limitation of PLCC links is their unreliability due to power line failures. This is caused by the open circuit problem. Situations such as open switches or disconnected power lines caused by disruptive events in the physical grid will lead to the loss of communication in certain parts of the grid, which results in less observability and controllability by the CC. This in turn makes it more difficult for the CC to halt the propagation of failure in the physical grid, which can then cause more failures in the PLCC links and hence further loss of observability and controllability. To break this harmful cycle without incurring exorbitant costs, it is crucial to judiciously combine PLCC and more reliable (and expensive) non-PLCC links in constructing the SCADA system, which is the goal of this work.

Summary of Contributions: Our main contributions are as follows:

- 1) In designing the communication network, we intend to mix the cost efficiency of PLCC links with the reliability of non-PLCC links. The optimization problem is formulated as maximizing an objective capturing the controllability of nodes in the grid, which is weighted by the node's importance in system topology and its generation/load contribution. The solution is a set of communication links, either PLCC or non-PLCC, in a coupled power grid with a geographically co-located

V. Farhadi, T. He, N. R. Chaudhuri, and T. LaPorta are with Pennsylvania State University, University Park, PA, 16802, USA (email: {vuf8, tzh58, nuc88, tfl12}@psu.edu). S. G. Vennelaganti is with Electric Power Research Institute, Palo Alto, CA, 94304, USA (email: svennelaganti@epri.com).

This work was supported by the National Science Foundation under award ECCS-1836827.

SCADA-based communication network, albeit with a budget constraint limiting the number and type of links (either PLCC or non-PLCC). We formulate the underlying optimization as non-linear programming (NLP).

2) By analyzing the complexity of our NLP problem, we prove that it is generally NP-hard. We derive a constant-factor approximation guarantee for the proposed formulation, computable by mixed-integer linear programming (MILP).

3) We perform extensive evaluations on IEEE 118-bus power system. The proposed algorithm consistently outperforms baselines while it effectively (i) estimates system topology in the absence of fully known $\mathbf{B}$ -matrix, (ii) increases the demand served after cascade.

Roadmap: Section III presents models of cascading failures and preventive control in the coupled system of a power grid and communication network. Section IV formulates the optimization problem and analyzes the complexity. Then, section V presents our proposed approximation algorithm and its performance guarantee. Section VI evaluates the performance of the proposed solution against benchmarks. Eventually, section VII concludes the paper.

II. BACKGROUND AND RELATED WORK

A. Background

Cascading failure – causes and models: There have been several instances of cascading failures in interdependent systems in recent decades [17]–[20], such as the 2003 blackout in Italy that severely impacted financial systems, the railway system, and telecommunication networks [19]. In this blackout, unanticipated problems in the communications both between the operators and towards the remote substation controls was one of the major reasons for the failures (especially during restoration) [19].

In general, interdependence tends to happen because of cross-system functional dependencies and geographical topology similarities [21]–[23]. These two factors imply the possible influences each network can have on the other in the state of cascading failures. For example, in the 2003 blackout in Italy, the shutdown of power stations initially caused losing some corresponding communication nodes and hence some information. This, in turn, made more power stations go into blackout [21]. Another example is the Northeast American blackout of 2003, in which a software bug originated a similar disaster [21]. Study [2] notes that blackouts in the US cost billions of dollars, and the losses due to blackouts increase dramatically with blackout duration. These examples show how inter-connectivity can significantly increase the scope and intensify the damage in an interdependent system during a large-scale cascade. For power grids, one solution is real-time monitoring and proper load shedding in the system.

It is important to be able to accurately model cascading failures in the power grid to develop and evaluate solutions for preventing them. There are three well-accepted cascading failure models of power systems: DC-quasi-steady-state (QSS), AC-QSS, and dynamic. AC-QSS models are based on AC power flow, which can capture voltage collapse in addition to

line overloading. Usually, these models suffer from divergence issues due to voltage collapse [24], [25]. The dynamic models [26], [27] present the most precise mechanism of cascade propagation. However, they are computationally costly for large-scale networks.

DC-QSS models, which we use in this paper, work based on DC power flow and are computationally economical and easily implementable [27]–[32]. These advantages offer the chance to build interdependent models of power and communication control networks and apply statistical analysis on a large-scale system. These models assume a uniform voltage profile and neglect the resistive loss; therefore, they cannot consider reactive power in the power system.

Communication links – PLCC vs. non-PLCC: PLCC links provide a means of supporting the SCADA system at a lower cost than using dedicated communication links, but PLCC has its own drawbacks. Power lines are primarily designed to be a transmission medium for electrical energy; hence, they may not be as suitable as data network media in terms of reliability, controllability, and security for data communication applications. Due to the complexity of transmitting data reliably over power lines, such systems also incur a nontrivial cost, albeit typically less than installing dedicated physically-separate communication media.

Due to the transmission properties of power lines, transmitted signals tend to attenuate over distance. Noise is added to the system by various loads and switching devices. PLCC links also radiate signals so the PLCC link itself has to meet current EMC (Electromagnetic compatibility) regulation limits to control radio signal interference [33]. To overcome the losses resulting from physical characteristics of power lines, repeaters are used to transmit signals over long distances, which re-amplify and re-package the signal to restore its previous signal level [34]. PLCC technology is being deployed mainly in the MV (medium voltage) distribution substations that have more suitable communication channels due to lower attenuation, time-invariant behavior, simplified network configurations without any branches, and fewer noise sources [35]. In MV topologies, the positions of repeaters are fixed at distribution substations [35].

Since power lines are already deployed in power grid, deployment costs of PLCC links are largely confined to connecting repeaters and modems to the existing electrical grids. In this work, for the sake of simplicity, we consider the average cost of placing PLCC links per meter as calculated in [15]. For non-PLCC links, fiber, copper, and microwave have been considered. Among these non-PLCC technologies, microwave provides the best flexibility and cost savings [15].

Admittance matrix for preventive control: The CC requires knowledge of the admittance matrix of the power grid to run the preventive control. In a power system with N buses, where each bus is connected to the other buses through transmission lines, an $N \times N$ admittance matrix (a.k.a. $\mathbf{B}$ -matrix) describes the nodal admittances of the various buses. Admittance, expressed in the unit "mho" (Ω), is the reciprocal of impedance, which is a complex number that measures how easily an element will allow a current to flow. Note that the impedance of a circuit element is the ratio of the phasor voltage

across the element to the phasor current through the element [36]. Most prior works in the domain of coupled power and communication networks assume constant availability of the $\mathbf{B}$ -matrix [30], [32]. In practice, the $\mathbf{B}$ -matrix is completely known before the cascade but only partially known as the cascade propagates. This is because some buses may lose their connectivity to the CC if the power lines that fail are used to implement PLCC links, the failure of which causes these buses and their incident lines to become unobservable to the CC.

This paper presents a procedure to design the control communication network of a power grid under a budget constraint, using an objective that represents the expected observability/controllability of the grid (to the CC) after initial failure. Our empirical evaluation shows that the proposed design effectively improves the accuracy of $\mathbf{B}$ -matrix estimation as well as the demand served after cascade.

Optimization for resource allocation: There are several well-known techniques for solving resource allocation problems. The most widely used optimization technique is linear programming (LP) [37]. However, many problems of interest cannot be exactly formulated as LP, in which cases researchers often attempt to approximate the non-linear constraints/objective function of their problem by linear functions and then solve the modified problem. Mixed-integer linear programming (MILP) is a generalization of LP such that some of the variables are constrained to be integers, while other variables are allowed to be non-integers. MILP is often used to solve large and complex optimization problems, as it balances imprecision from linearization while taking advantage of the well-defined global optima and efficient commercial-grade solvers that are generally unavailable for mixed-integer non-linear programming formulations [38]–[40].

This work develops an approximation solution to the proposed resource allocation problem by relaxing the original non-linear programming problem into a MILP (see Section V).

B. Related Work

Despite notable accomplishments in the study of power grid cascading failures, cascades in a coupled power grid system and associated communication network are not well studied. Authors in [21], [30]–[32], [41] assume a co-located power-communication overlay. Reference [29] shows that the DC-QSS power grid model captures important information regarding the development and final impact of cascading failures. In [30], authors compare a topological contagion model to a DC-QSS power grid model and conclude that the system’s vulnerability decreases by improving the power-communication coupling. Authors in [41] came to the same conclusion. Reference [31] uses the DC-QSS model for the power grid and assumes that the location of failures in a coupled power-communication network might be unknown. The authors propose a coupled power-communication model such that the operation of its monitoring system and performance of the power grid are mutually dependant. The work closest to ours is [32], which assumes that all lines are equipped with PLCC links. It proposes algorithms to select a set of PLCC lines to install non-PLCC

links with a budget constraint on such links. Moreover, it deals with an idealistic assumption that all lines’ breaker status, i.e., the $\mathbf{B}$ -matrix, are known. Therefore, the resource allocation focus is solely on maximizing the controllability of generators and loads, given the budget constraint of non-PLCC links.

This paper also examines the DC-QSS model of cascading failures in a power grid coupled with a SCADA-based communication network as in [32], with the following key differences:

- 1) Unlike [32], the ideological assumption of the known $\mathbf{B}$ -matrix is relaxed, so we rely solely on the status of observable breakers. The new assumption poses new challenges, e.g., the preventive control requires estimating the $\mathbf{B}$ -matrix. It also realistically captures the impact of the loss of observations due to communication links failing because without full sensor readings, the $\mathbf{B}$ -matrix cannot be fully reconstructed. The estimation of the admittance matrix can be divided into two parts; first the islands formed after failures must be detected, and second the connectivity of unobservable lines/nodes within islands must be estimated. The effectiveness of the preventive control depends upon the accuracy of both.
- 2) While ensuring the observability of all nodes at the CC before cascade, we consider the total cost of all communication technologies, including both PLCC and non-PLCC links. The reason is that deploying PLCC links also incurs a cost, and it is not realistic to assume all links are PLCC.

C. Gap and Challenges

Literature	Dependency	Comm. link
[29], [42], [43]	none (standalone grid)	-
[44]–[46]	one-way (control $\rightarrow$ grid)	non-PLCC
[21]–[23], [30], [31], [41]	two-way (control $\leftrightarrow$ grid)	non-PLCC
[32]	two-way (control $\leftrightarrow$ grid)	PLCC & non-PLCC

Table I: Literature on cascading failure in power grid in presence/absence of control network

Table I summarizes literature on cascading failure in the power grid in presence/absence of control network, indicating the interaction between the power grid and control network as well as the authors’ assumption about the type of communication links. Real-world examples [2], [17], [19] emphasize the significance of understanding the interdependency between the coupled systems of the power grid and communication network to prevent cascading failures from causing massive blackouts. The majority of the existing works focus on the interdependency between the control network and the power system under the simplistic assumption of secure and full connectivity between all the nodes and the control center via non-PLCC links such as fiber optic links. The important questions of “*how to build that secure connectivity*” and “*what will happen in the absence of the secure and full connectivity*” of the control network have remained unanswered. Our work addresses this gap by

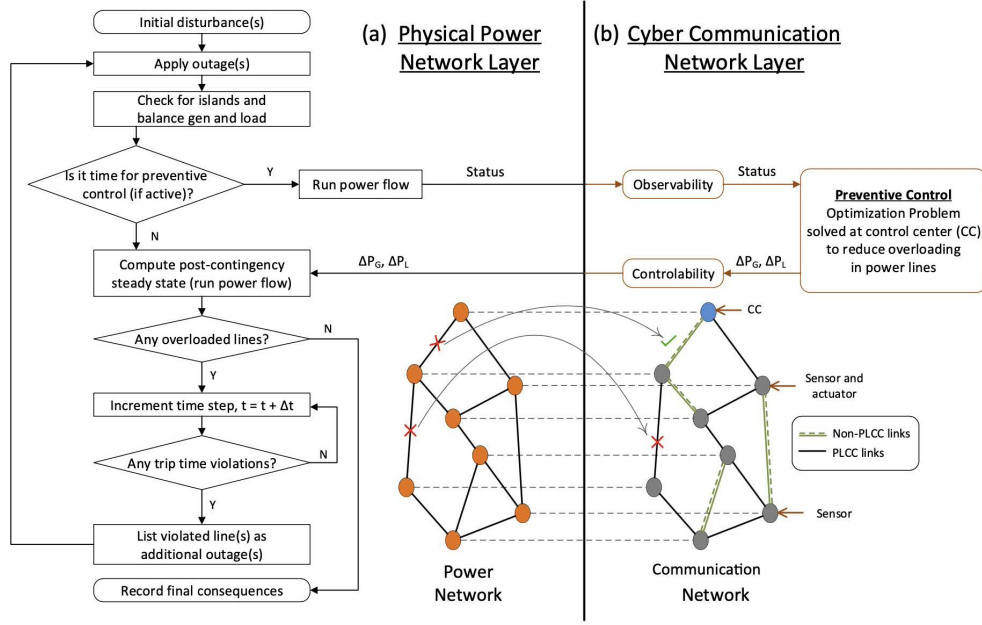


Figure 1: Flow chart highlighting the modeling of the coupled cascading failure in power and communication networks.

characterizing the impact of having a realistic control network coupled to the power grid on the algorithms for detecting and stopping cascades, and optimizing the tradeoff between the reliability and the cost of constructing such a control network.

To the best of our knowledge, we are the first to study the design of control networks comprised of PLCC and non-PLCC links while considering the communication needs during cascade mitigation. The main challenges in solving this problem are: (i) the effect of control network design on the propagation of cascade is highly non-linear and non-explicit, and (ii) the solution space for the design is discrete and can be very large for large grids. In this work, we present an optimization-based approach to tackle both challenges that is shown to effectively support control algorithms in topology estimation and preventive control in the face of cascading failure.

III. SYSTEM MODEL & MOTIVATION

This paper studies cascading failure in a coupled power grid with a geographically co-located SCADA-based communication network. In summary, a Control Center (CC) gathers data from sensors/actuators and dispatches preventive control commands to generators and loads. This section provides specifics of the cascading failure and preventive control model of the power grid.

A. System Model

In this work, we model the power grid as a undirected graph $G = (V, E)$ with no self-loops or multiple edges, where the V and E are the buses and the transmission lines, respectively. We also consider a geographically co-located SCADA-based communication network, albeit with a budget constraint limiting the number and type of links. Fig. 1 shows the interaction between the power system and the communication network during cascade. After applying initial disturbance-based outages, we detect islands in the power grid and balance the generation and load in each of them. The

balance is obtained either by curtailing the generation across all generators or reducing the load over all load nodes uniformly. A complete blackout in an island results from the lack of at least one active generator within the island.

Over-current relays release tripping commands when a branch remains overloaded until its trip time elapses. The trip time is computed based on the line's standard inverse-time over-current characteristics. This means that all the potential line trips do not necessarily end up being tripped. A settled post-contingency steady state is obtained by repeating the entire process until no potential line trip is left.

If the preventive controller is activated, as shown in Fig. 1, all the known power grid information, including line flows, the status of breakers, power output of generators and consumption at load centers, are communicated through communication links, either by power lines (PLCC links) or dedicated communication infrastructures (non-PLCC links). Once all the relevant information is passed on to the preventive controller, the CC tries to reduce line overloading by issuing control commands, which are communicated back to the power grid.

B. Preventive Control & Topology Estimation

To prevent cascade propagation, centralized generation redispatch and load shedding can be deployed using the control network. The admittance matrix ($\mathbf{B}$ -matrix) provides the topology of the power grid, as well as the information needed for the load/generation reduction or the power flow study of buses. Although the complete system information is available before the outage, it is only partially known as the cascade proceeds. Due to interdependency between cyber and physical layers of the grid, failure in the power grid during a cascade leads to outages in the communication network, which gradually decreases the observable areas. Hence, in the absence of a fully known admittance matrix, we need to estimate it. In the following, we focus on the preventive control and estimation of the power grid admittance matrix ($\hat{\mathbf{B}}$) in detail.

1) *Example Application 1 - Preventive Control*: The CC houses the preventive controller in the geographically co-located SCADA-based communication network. The CC is connected to all sensors and actuators and has the updated information of all the elements in the power grid. It controls the propagation of the failures in the power grid by reducing the overloading in power lines. It is important to note that failures in the power grid may significantly affect the connections between the CC and these sensors/actuators if PLCC communication is used and no backup communication line is applied. Therefore, cascading failures in the power grid may also lead to failures in the communication network, which results in more sensors becoming unobservable and more actuators becoming uncontrollable by the CC. This issue, in turn, deteriorates the effectiveness of cascade prevention.

The CC, by running the following optimization problem (1) [30] at regular intervals, minimizes the sum of total amount of load shedding ($-\mathbf{1}^\top \Delta \mathbf{P}_L$) and weighted overloads ($\lambda^\top \mathbf{L}_{over}$) in the power grid. The inputs and output of formulation (1) are line status and branch flows ($\mathbf{L}_{\mathcal{M}}$) of the measurable set of buses ($\mathcal{M}$), and load shedding ($\Delta \mathbf{P}_L$) and generation reduction ($\Delta \mathbf{P}_G$) values, respectively. In this formulation, $\bar{\mathcal{M}}$ is the unmeasurable set of buses. $\mathbf{P}_L$ and $\mathbf{P}_G$ show the load and generation power at each bus, θ_i presents the phase angle at the i th bus, and λ is the uniform weight vector. The variables $\mathbf{L}$, $\mathbf{L}_{max}$ and $\mathbf{L}_{over}$ represent the actual power flow, its allowable maximum value, and overloading, respectively. The principal used notations are explained in Table IV.

$$\min_{\Delta \mathbf{P}_G, \Delta \mathbf{P}_L} -\mathbf{1}^\top \Delta \mathbf{P}_L + \lambda^\top \mathbf{L}_{over} \quad (1a)$$

$$\text{s.t. } \Delta \mathbf{P}_G - \Delta \mathbf{P}_L = \mathbf{B} \Delta \theta, \quad \Delta \theta_i = 0, \forall i \in \Omega_{ref} \quad (1b)$$

$$\Delta L_{ij} = \frac{\Delta \theta_i - \Delta \theta_j}{x_{ij}}, \quad \forall i, j \in \mathcal{M} \quad (1c)$$

$$|\mathbf{L}_{\mathcal{M}} + \Delta \mathbf{L}| \leq \mathbf{L}_{max} + \mathbf{L}_{over}, \quad \mathbf{L}_{over} \geq \mathbf{0} \quad (1d)$$

$$-(\mathbf{P}_G)_{\mathcal{M}} \leq (\Delta \mathbf{P}_G)_{\mathcal{M}} \leq \mathbf{0} \quad (1e)$$

$$-(\mathbf{P}_L)_{\mathcal{M}} \leq (\Delta \mathbf{P}_L)_{\mathcal{M}} \leq \mathbf{0} \quad (1f)$$

$$(\Delta \mathbf{P}_G)_{\bar{\mathcal{M}}} = \mathbf{0}, (\Delta \mathbf{P}_L)_{\bar{\mathcal{M}}} = \mathbf{0} \quad (1g)$$

2) *Example Application 2 - Topology Estimation*: The admittance matrix is of critical importance because it is used to analyze the data needed in the load/generation reduction or the power flow study of buses; for example, breaker statuses indicate tripped lines in the power system. In summary, the $\mathbf{B}$ -matrix explains the admittance and the topology of the power system. In the absence of a fully known $\mathbf{B}$ -matrix, formulation (1) takes $\hat{\mathbf{B}}$ (the estimated admittance matrix of the power grid), proposed in [47] as input, such that (1b) is replaced with

$$\Delta \mathbf{P}_G - \Delta \mathbf{P}_L = \hat{\mathbf{B}} \Delta \theta, \quad \Delta \theta_i = 0, \forall i \in \Omega_{ref} \quad (2a)$$

As the cascade propagates, multiple line outages may happen in sequence, which leads to the formation of many islands within the power grid. Authors in [47] first identify the buses forming the connected components and then detect further

line outages within the individual islands. Their topology identification algorithm uses power system measurements and observable breaker statuses. Also, their proposed estimation is verified at every step of the cascading failure. As the estimation methodology of the admittance matrix is outside the scope of this paper, we skip the detailed explanation of it and refer the audience to [47]. We use this method of estimating the admittance matrix in preventive control of cascading failure formulated (1), in a closed-loop fashion.

The principle of energy conservation in each island requires the existence of at least one active generator to produce power and at least one non-zero load to consume the power. The island will be dead if there is no active generator or non-zero load within it. As this paper focuses on the re-dispatch control sub-problem, it necessitates an accurate island estimation. To that end, at least one generator and one load must be observable within that particular island. The results in [47] show that the identification is accurate in more than 95% of cases when at least 50% of nodes are observable in each island (including at least one generator and one load bus). Hence, for the purpose of preventive control, we focus on ensuring observability of generator and load nodes from the CC's points of view that can provide these two objectives: (i) observing at least one active generator and one non-zero load in each island and (ii) observing at least 50% of nodes in each island. Satisfying these objectives helps identify the islands, followed by identifying the $\hat{\mathbf{B}}$ within each island.

These two focus areas are complementary, but neither is solely enough to ensure maximum served demand. The former approaches the cascade prevention purely from a controllability standpoint, while the latter considers the sub-problem from the observability side. For the sake of simplicity, we refer to islands consisting of at least one active generator and one non-zero load as valuable islands. Highly observable islands are valuable islands with at least 50% nodes observable by the CC.

C. Motivating Experiment

To understand the potential value of preventive control in mitigating cascading failures, we evaluate the performance of a network with preventive control. Considering the IEEE 118-bus system [48] as a test system, which includes 118 buses, 186 branches, and 54 generators. We randomly fail 5% of the buses in the power grid and calculate the served power after

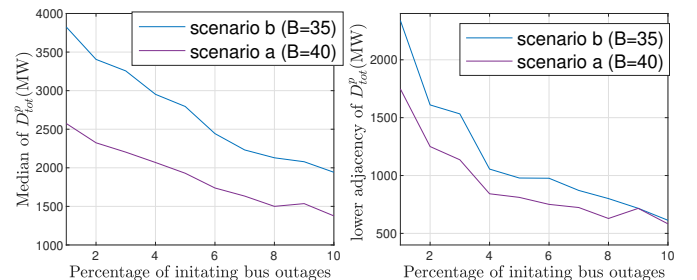


Figure 2: Performance evaluation of scenarios (a) randomly selecting PLCC links under $B = 40$ and (b) mixture of suitably placed PLCC and non-PLCC links under $B = 35$, in terms of median and lower adjacency of D_{tot}^P .

cascade propagation in scenarios (i) having no communication network versus (ii) a network of 100%-PLCC links and (iii) a network of 100%-non-PLCC links. In scenarios (ii) and (iii), the communication networks are geographically co-located with the power grids and have the same topology as power grids. Assuming each PLCC link costs 0.3 and non-PLCC costs 0.6, the total cost for the communication network is $\mathcal{B} = 55.8$ and $\mathcal{B} = 111.6$ for scenarios (ii) and (iii), respectively. In this experiment, 100 random cases were tested. Results show that the mean of residual power after cascade propagation are 2041.3 MW (47.57% of the initial power), 2492.1 MW (58.08% of the initial power), and 3038.3 MW (70.81% of the initial power) for scenarios (i), (ii), and (iii), respectively. As it was expected superior performance is achieved with a pure non-PLCC communication network, albeit at a higher cost.

In the next example, we show that by strategically-placing non-PLCC and PLCC links, we achieve better efficacy in mitigating cascading failures compared with a more expensive communication network that only employs PLCC links. We describe the algorithms we use to determine this placement in the remainder of the paper; here we are just summarizing the large benefit of a good solution. We randomly fail 1%-10% of the buses in the power grid and calculate the served power after cascade propagation in scenarios (a) having a network of randomly placed PLCC links under budget $\mathcal{B} = 40$ versus (b) communication network that is mixture of appropriately placed PLCC and non-PLCC links under budget $\mathcal{B} = 35$. In this experiment, 100 random cases were tested. Results in Fig.2 show superior performance in the latter scenario, although its budget is tighter.

In support of topology estimation and the accuracy of the $\hat{\mathbf{B}}$ (the estimated admittance matrix of the power grid), Table II and Table III demonstrate the importance of installing the appropriate type of links in the communication network. Table II compares the percentage of valuable and highly observable islands, where valuable islands consist of at least one active generator and one non-zero load. Highly observable islands are the valuable islands with at least 50% nodes observable by the CC, otherwise they are considered as slightly observable islands (the valuable islands with less than 50% nodes observable by the CC). Furthermore, Table III checks the percentage of correctly identified islands.

These examples demonstrate the potential value of preventive control and the role of types of links in mitigating cascading

failures, jointly considering the budget.

IV. PROBLEM FORMULATION

Table IV: Table of notations

Notation	meaning
$\mathbf{P}_G$	generation power at each bus
$\mathbf{P}_L$	load power at each bus
$\mathbf{L}$	actual power flow
$\mathbf{L}_{over}$	overloading in lines
$\mathcal{M} / \bar{\mathcal{M}}$	measurable / unmeasurable set of buses
λ	uniform weight vector
$\Delta\theta_i$	change in the phase angle at the i th bus
$\mathbf{B}$	admittance matrix
$\hat{\mathbf{B}}$	estimated admittance matrix
V	set of nodes in the power grid
E	set of links in the power grid
$\mathcal{C}$	set of communication link platforms
w_c	cost of communication link type c
s_e^c	binary variable shows type c of communication link e
f_{ij}^k	flow of commodity k from node i to node j
$\mathcal{B}$	communication links budget
P_i	the set of all possible paths between the CC and node i
γ_i	the importance of observing/controlling node i by the CC
p_c	reliability of communication link type c

Without a properly designed communication network, a SCADA-based system cannot work adequately. All supervisory control and data acquisition aspects of the SCADA system rely entirely on the communication system to provide a conduit for data flow.

This paper's main improvement in providing sufficient observability and controllability for cascade prevention is observing and mitigating cascades in valuable islands. Also, to guarantee the CC's communication with these islands, it is crucial to correctly place PLCC and non-PLCC links, as failure in the power line fails the PLCC link piggybacked on the line. Moreover, we assume all the communication nodes, i.e., sensors, actuators, and relays, have battery backups and are immune from power grid failures.

A. Underlying Optimization Problem

We formulate the problem as link placement optimization in the communication network, considering both PLCC and non-PLCC links, albeit under a budget of $\mathcal{B}$ to capture the resource constraints. In this way, we combine the cost efficiency of PLCC links and the robustness against cascades of non-PLCC links. Although cascade prevention aims to maximize the total demand served after the cascade, this objective function is not an explicit function of link placement. To address this challenge, we propose using an objective function as follows.

Without loss of generality, we assume that the given power grid topology $G = (V, E)$ is a undirected graph with no self-loops or multiple edges. Let $P_i, \forall i \in V$ denote the set of all

Table II: Percentage of (i) valuable and (ii) highly observable islands in different scenarios (a) randomly selecting PLCC links under $\mathcal{B} = 40$ and (b) mixture of suitably placed PLCC and non-PLCC links under $\mathcal{B} = 35$.

	scenario (a)	scenario (b)
(i)	16.38	32.97
(ii)	3.01	26.34

Table III: Percentage of estimation accuracy in different scenarios (a) randomly selecting PLCC links under $\mathcal{B} = 40$ and (b) mixture of suitably placed PLCC and non-PLCC links under $\mathcal{B} = 35$ for (i) valuable, (ii) highly observable and (iii) slightly observable islands.

	scenario (a)	scenario (b)
(i)	39.54	87.48
(ii)	51.67	93.55
(iii)	33.11	52.20

possible paths between the CC and node i . By modeling each path $m \in P_i$ as a set of traversing links, the total reliability of the network is defined as follows [49] (assuming nodes never fail), where ρ_e denotes the reliability of link e .

$$\sum_{i \in V} \left(1 - \prod_{m \in P_i} (1 - \prod_{e \in m} \rho_e) \right), \quad (3)$$

Let the binary variables $s_e^c \in \{0, 1\}$ indicate selection of link e by type $c \in \mathcal{C}$, and p_c denote the reliability of type c link (PLCC or non-PLCC). Reliability of link e is defined as $\rho_e := \sum_{c \in \mathcal{C}} s_e^c p_c$. Thus, the link placement problem (4) maximizes the total reliability of the network under a given budget of PLCC and non-PLCC links, while ensuring each node is observable by the CC. The principal used notations are described in Table IV.

$$\max \sum_{i \in V} \gamma_i \left(1 - \prod_{m \in P_i} (1 - \prod_{e \in m} \sum_{c \in \mathcal{C}} s_e^c p_c) \right) \quad (4a)$$

$$\text{s.t. } \sum_{e \in E} \sum_{c \in \mathcal{C}} s_e^c w_c \leq \mathcal{B} \quad (4b)$$

$$\sum_{c \in \mathcal{C}} s_e^c \leq 1, \forall e \in E \quad (4c)$$

$$\sum_{j \in V} f_{ij}^k - \sum_{j \in V} f_{ji}^k = \begin{cases} 1 & i = CC \\ -1 & i = k \\ 0 & \text{otherwise} \end{cases} \quad (4d)$$

$$\forall i \in V, \forall k \in V \setminus \{CC\}$$

$$f_{ij}^k \leq \sum_{c \in \mathcal{C}} s_e^c, \forall e = \{(i, j) \vee (j, i)\} \in E, k \in V \setminus \{CC\} \quad (4e)$$

$$f_{ij}^k \geq 0, \forall e = \{(i, j) \vee (j, i)\} \in E, k \in V \setminus \{CC\} \quad (4f)$$

$$s_e^c \in \{0, 1\}, \forall e \in E, \forall c \in \mathcal{C} \quad (4g)$$

Here, binary variables $s_e^c \in \{0, 1\}$ indicate the selection of type c for link e and the variable f_{ij}^k for each arc (i, j) symbolizes the flow of commodity k from node i to node j . Briefly, the Non-Linear Programming (NLP) (4) aims at selecting PLCC and non-PLCC links up to budget $\mathcal{B}$ to maximize the expected total weight of all the nodes remaining connected to the CC after the initial failure. Intuitively, the more nodes connected to the CC, the better the CC observes and controls the grid, and hence mitigates the failure cascading properly. However, not all the nodes are equally important; therefore, we use the weight γ_i to consider the importance of observing and controlling node i . Constraints (4b) and (4c) make sure that the budget is not exceeded, and for each link at most one type of communication link is selected, respectively. Constraints (4d)-(4f) guaranty the observability of each node by the CC before any failure occurs, which is based on the link constrained Steiner tree problem in undirected graphs, described in the following.

Definition IV.1 ([50]). *Given the undirected graph $G(V, E)$ with non-negative edge cost, Link Constrained Steiner Tree problem of V determines the Steiner tree $T = (V, E(T))$ rooted*

at a source node with minimum cost and such that the number of edges $E(T) \in E$ is less than or equal to a given threshold.

Graph construction: Given the undirected graph $G(V, E)$, construct the bi-directed graph $\mathcal{H} = (V, \mathcal{A})$ such that an edge $e \in E$ incident to nodes i and j , is replaced by two arcs, that is, $\{(i, j), (j, i)\} \in \mathcal{A}$. Considering a commodity k for each node $V \setminus \{CC\}$, the variable f_{ij}^k for each arc $(i, j) \in \mathcal{A}$, represents the flow of commodity k from node i to node j .

Claim: The solution of link constraint Steiner tree problem on graph $\mathcal{H} = (V, \mathcal{A})$ provides a solution to the constraints (4d)-(4f) on graph $G = (V, E)$.

Proof of the claim: The Steiner tree $T = (V, E(T))$ of graph $\mathcal{H}$ obtains a Steiner arborescence rooted at node CC, containing a directed path from node CC to every other terminal node in $V \setminus \{CC\}$. The variable s_e^c takes value equal to 1 if corresponding edge $e \in E(T)$ and 0 otherwise.

B. Complexity analysis

Consider the special case that the costs of different communication links are $\{1, 0\}$ for non-PLCC and PLCC links, respectively. Since PLCC costs nothing, without loss of generality we can assume that $\sum_{c \in \mathcal{C}} s_e^c = 1$ for every link e , thus constraint (4c) is unnecessary. Moreover, constraints (4d)-(4f) are no longer needed as there is a communication path between the CC and every other node (assuming the power grid topology is connected before failure through PLCC links). Thus, the NLP formulation (4) changes to (5) by keeping equations (4a), (4b) and (4g).

$$\max \sum_{i \in V} \gamma_i \left(1 - \prod_{m \in P_i} (1 - \prod_{e \in m} \sum_{c \in \mathcal{C}} s_e^c p_c) \right) \quad (5a)$$

$$\text{s.t. } \sum_{e \in E} \sum_{c \in \mathcal{C}} s_e^c \leq \mathcal{B} \quad (5b)$$

$$s_e^c \in \{0, 1\} \quad \forall e \in E, \forall c \in \mathcal{C} \quad (5c)$$

Theorem 1. *Problem (5) is NP-hard.*

Proof. Formulation (4) is a generalization of the NP-hard optimization in [32]. $\square$

NP-hardness of the special case (5), proves (4) is NP-hard too.

V. APPROXIMATION ALGORITHM

In the previous section, we proposed the link placement problem (4), which maximizes the total reliability of the network from the CC's point of view, based on jointly considering the topology and power system information metrics. In this section, we explore an efficient solvable solution, which derives a constant-factor approximation algorithm for the formulation (4) that is computable by MILP.

Optimization link placement problem (4) aims at maximizing (4a), which is the root of non-linearity. Let $p_{\min} := \min_{c \in \mathcal{C}} p_c$ and $p_{\max} := \max_{c \in \mathcal{C}} p_c$ denote the minimum and

maximum reliability per link, where p_c shows the reliability of link type c .

Assume $V = V_1 \cup V_2 \cup V \setminus \{V_1 \cup V_2\}$ such that $V_1 := \{i \in V : \sum_{m \in P_i} \prod_{l \in m} p_{max} \leq 1\}$ and $V_2 := \{i \in V : \prod_{l \in m_{2,max}} p_{min} \geq 1\}$ can be any subsets of V , where $|m_{2,max}|$ is the maximum hop count per path over all the paths in $\bigcup_{i \in V_2} P_i$. For the sake of simplicity, let $A_l = \sum_{c \in \mathcal{C}} s_l^c p_c$. Hence, the objective function (4a) changes to (6):

$$\begin{aligned} & \max \sum_{i \in V_1} \gamma_i \left(1 - \prod_{m \in P_i} (1 - \prod_{l \in m} A_l) \right) + \\ & \sum_{i \in V_2} \gamma_i \left(1 - \prod_{m \in P_i} (1 - \prod_{l \in m} A_l) \right) + \\ & \sum_{i \in V \setminus \{V_1 \cup V_2\}} \gamma_i \left(1 - \prod_{m \in P_i} (1 - \prod_{l \in m} A_l) \right) \end{aligned} \quad (6)$$

Theorem 2. *The optimal solution to (7) yields a $(1 - \epsilon)$ -approximation for (4) under the condition of $p_{min}^{|m_{3,max}|} \geq p_{max}^{|m_{3,min}|} (1 - \frac{1}{e})^{|P_{max}|}$, where $\epsilon := (e - (e - 1)(1 - p_{min})^{|m_{1,max}|})^{-1}$ such that $|m_{1,max}|$ denotes the maximum hop count per path over all the paths in $\bigcup_{i \in V_1} P_i$. Moreover, $|m_{3,min}|$ defines the minimum hop count per path over all the paths in $\bigcup_{i \in V \setminus \{V_1 \cup V_2\}} P_i$ and $|P_{max}| := \max_{i \in V \setminus \{V_1 \cup V_2\}} |P_i|$.*

$$\begin{aligned} & \max UB + UB' + UB'' \\ & \text{s.t. (4b) - (4g),} \end{aligned} \quad (7)$$

where UB , UB' and UB'' are the upper-bounds of the first, the second, and the third terms of (6), respectively, defined as:

$$UB := \sum_{i \in V_1} \gamma_i \sum_{m \in P_i} \left(\left(1 - \left(1 - \frac{1}{e} \right) \sum_{l \in m} (1 - A_l) \right) \right) \quad (8)$$

$$UB' := \sum_{i \in V_2} \gamma_i \quad (9)$$

$$UB'' := \sum_{i \in V \setminus \{V_1 \cup V_2\}} \gamma_i \left(1 - \prod_{m \in P_i} (1 - \prod_{l \in m} p_{max}) \right) \quad (10)$$

Please refer to appendix A for the proof.

VI. PERFORMANCE EVALUATION

This section demonstrates the performance of the proposed approximation algorithm on an electric power system.

A. Benchmarks and metrics

To assess the performance of the proposed algorithm, we use the following benchmarks:

- *The approximation algorithm* (7), using a MILP solver (MATLAB `intlinprog`);
- *LP-relaxation with rounding*, which first solves the LP-relaxation of proposed approximation algorithm (7), and

then rounds the link selection variables to $\{0, 1\}$, subject to constraints (4b)-(4c);

- $(1 + \epsilon, 1)$ -approximation algorithm of CMST, which is explained in part VI-A1 in the following;
- $(1, 1 + \epsilon)$ -approximation algorithm of CMST, which is explained in part VI-A1 in the following;
- *BC method*, which is explained in part VI-A2 in the following;
- *Random*, which randomly selects links and their type until exhausting the budget or assuring the CC's connectivity to all nodes.

1) (α, β) -Approximation of CMST: The general NP-hardness of MILP approximation algorithm (7) motivates us to develop the following alternative solution by focusing solely on the graph-theoretic metric of the network. In this method, we form a special case of the constrained minimum spanning tree (CMST) problem based on the power grid topology. Then, we apply an (α, β) -approximation of it, albeit with a budget constraint. Although this method differs from (4) in the objective function, we will show that it is polynomial-time solvable while achieving considerably better performance compared to random in terms of served loads and accuracy of $\hat{\mathbf{B}}$. Recalling that the solution of (4) is a spanning tree, the output of this method is also a spanning tree. This method reveals that having a connected communication network such that the CC is connected to all nodes, does not necessarily guarantee the best performance, as shown later in Fig. 8.

Let $\mathcal{H}' = (V, \mathcal{A}')$ be a undirected graph and W is a positive integer. Assume two different non-negative functions, weight and cost, associate on edges in $\mathcal{A}'$.

Definition VI.1 ([51]). *The constrained minimum spanning tree (CMST) problem on graph $\mathcal{H}'$ is to identify a minimum total cost spanning tree on graph $\mathcal{H}'$, such that the total weight is at most W .*

Definition VI.2 ([51]). *Given two positive real numbers α and β , an (α, β) -approximation algorithm for the CMST problem on graph $\mathcal{H}'$ is defined as a polynomial-time algorithm that returns a solution with the total weight at most α times the bound W , and the total cost at most β times the total cost of the optimal solution for the CMST problem.*

Graph construction: Given the power grid topology $G(V, E)$, construct graph $\mathcal{H}' = (V, \mathcal{A}')$, such that an edge $e \in E$ is replaced by two edges with associated weights equal implementation cost of PLCC and non-PLCC link. Assign each edge in $\mathcal{A}'$ the cost equals $1 - p_c$, where p_c is the reliability of edge e with type c .

Claim: Assume W equals $\mathcal{B}$ (the budget of communication links), and $\epsilon > 0$ is a positive constant. Then, $(1, 1 + \epsilon)$ -approximation algorithm [51] and $(1 + \epsilon, 1)$ -approximation algorithm [52] of the CMST problem on graph $\mathcal{H}'$, provide $(1, 1 + \epsilon)$ and $(1 + \epsilon, 1)$ -approximation solutions of problem (4) on graph G , respectively.

Proof of the claim: The approximation solution $T' = (V, E(T'))$ on graph $\mathcal{H}' = (V, \mathcal{A}')$ obtains a

spanning tree rooted at node CC, containing a directed path from node CC to every other node in $V \setminus \{\text{CC}\}$. The variable s_e^c takes value equal to 1 if corresponding edge $e \in E(T')$ and 0 otherwise.

Complexity: Given undirected graph $\mathcal{H}' = (V, \mathcal{A}')$, the time complexity of two polynomial time $(1, 1 + \varepsilon)$ -approximation and $(1 + \varepsilon, 1)$ -approximation algorithms of CMST on graph $\mathcal{H}'$ are $O((|V| \log \log |V| + |V|^{1/((1+\varepsilon)^\zeta - 1)} \log \log(1 + \varepsilon))(|\mathcal{A}'| \log^2 |V| + |V| \log^3 |V|))$ [51], and $O(|V|^{O(\frac{1}{\varepsilon})} (|\mathcal{A}'| \log^2 |V| + |V| \log^3 |V|))$ [52] for any constant $\varepsilon > 0$, respectively. $\zeta > 0$ is a constant. As we construct graph $\mathcal{H}' = (V, \mathcal{A}')$ from the power grid topology $G(V, E)$, hence $|\mathcal{A}'| = 2|E|$.

2) *BC method:* Given the power grid topology $G(V, E)$, this method sequentially considers each link $e \in E$ and computes the value of BC_e equal to the maximum betweenness centrality of its two endpoints. The betweenness centrality of a node is the frequency that it appears on the shortest paths between all pairs of nodes in the graph [53]. Then, in the descending order of BC_e , it selects non-PLCC links until reaching 5% of the total budget (the proposed approximation algorithm (7) utilizing almost the same non-PLCC links), and PLCC links until exhausting the budget or assuring the CC's connectivity with all other nodes.

Remark: The BC method gives priority to links incident to nodes with higher betweenness centrality. Intuitively, these nodes have substantial influence over the information passing between other nodes. Thus, the goal is to maintain as many paths from these elements in the power grid to the CC.

To demonstrate the effectiveness of the proposed algorithm, we use two metrics supporting different applications: the ratio of valuable and highly observable islands in support of topology estimation, and the statistical measurements of load served after failure cascade in support of preventive control. Recall that valuable islands are islands consisting of at least one active generator and one non-zero load, and highly observable islands are the valuable islands with at least 50% nodes observable by the CC. In the former metric, the total numbers of formed islands are normalized by the 2-norm method [54]. For the latter metric, we show the lower adjacency and the median of the change in total post-contingency demand served (D_{tot}^p). The lower adjacency is the smallest data point that is not an outlier in the plot, which is 1st quartile - $1.5 \times$ inter-quartile range. D_{tot}^p before any failure is 4291 MW. Moreover, to examine the accuracy of the $\hat{\mathbf{B}}$, we check the percentage of correctly identified islands in some scenarios and eventually compare misses and false alarms in identifying the connected/tripped lines within the islands.

B. Simulation Setup

We evaluate the proposed solution on the IEEE 118-bus system [48], including 118 buses, 186 branches, and 54 generators. We study cases with initial bus outages varying from 1% – 10% of the total buses. For each case, 100 random

sets of node outages have been considered, such that all result in cascade. The CC is situated on bus 49, one of the highest degree nodes in the power grid system. The preventive control optimization and the delay in line tripping are both set to 80 seconds.

It is difficult to get data on the implementation cost of different types of communication links. In general, fiber and then microwave links have the highest implementation cost, while PLCC links are the cheapest [55]. Authors in [15] and [16] estimate the implementation cost of PLCC, microwave, and fiber as \$123, \$241 and \$450 per meter, respectively, where in this work, we normalize these costs. For the sake of simplicity, we consider all non-PLCC links as microwave, and set the cost of non-PLCC links twice the cost of PLCC links, which are 0.6 and 0.3, respectively. Note that this parameter setting implies that all links are assumed to have the same length, but our solution can be easily extended to incorporate heterogeneous lengths.

Further studies including implementing fiber and microwave together are left to future work. We set reliability, p_c , of PLCC and non-PLCC links to 0.99 and 0.9999 [56]. Therefore, equations (22) and (26) hold for special cases $V_2 = \{\emptyset\}$, $|P_{max}| = 1$ and $|m_{3,max}| \leq 45$.

C. Results

To provide a general understanding of different algorithms, we depict design outcomes of the benchmarks discussed in part VI-A in Fig. 3. Figure 3(a) shows the IEEE 118-bus power grid topology, and Figs. 3(b-g) offer the communication layers of the IEEE 118-bus power system under different designs before imposing any failure. The budget is set to 40. It is sensible to conclude that different link selection policies lead to picking various links as PLCC and non-PLCC, which further affect the performance of cascade prevention. Next, we will explain the settings used in these benchmarks to produce the corresponding communication networks in more detail.

1) Setting Design Parameters:

Comparison under node weight definition: Recalling that γ_i represents the importance of observing node i by the CC, we examine both the topological (centrality and degree) and the service (power injection) importance of node i . Moreover, we impose upper-bounds L and N on the length and number of paths between the CC and any particular node. This is justified because the throughput of a flow drops as the hop count increases.

We compare performance of the approximation algorithm (7) under four different definitions of weights: (i) power injection $\times$ BC (Betweenness centrality), (ii) BC, (iii) degree, and (iv) power injection of nodes solely. Upper-bounds $\mathcal{B} = 40$, $N = 5$ and $L = 20$ are also imposed.

The results in Fig. 4 and Table V show that the node weight definition of “the BC of the node $\times$ the real power injected at the node” attains the best performance in both the total load served after cascade and topology estimation, as it considers both the topological (BC) and the service (power injection) importance of node i . For the rest of the results in this section,

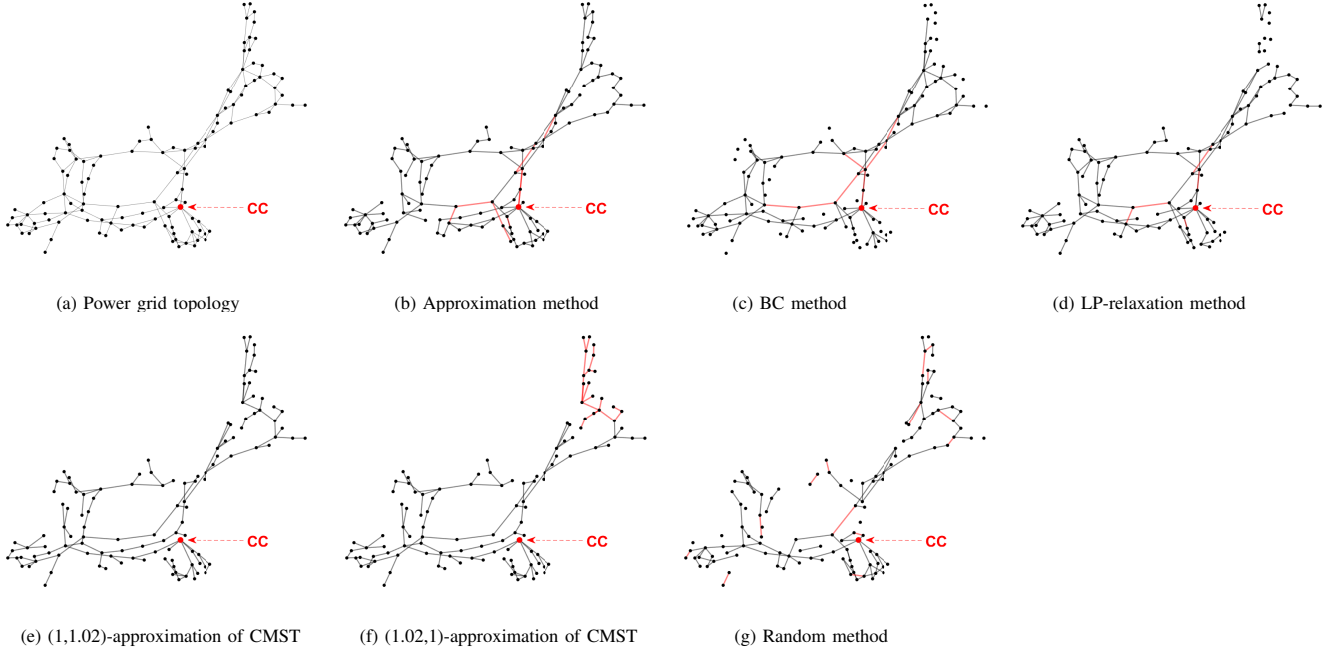


Figure 3: Graphs showing (a) the power grid topology and (b-g) the communication network obtained from different design methods. Black edges: PLCC, red edges: non-PLCC, CC: control center

we will use this definition of weight for the approximation algorithm.

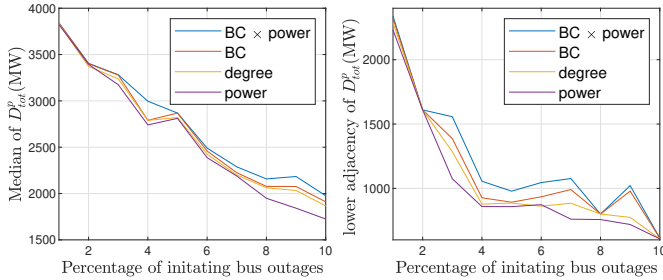


Figure 4: Performance evaluation of approximation algorithm in terms of median and lower adjacency of D_{tot}^P for different node weight definition under $B = 40$, $N = 5$ and $L = 20$.

Table V: The effect of different node weight definition on approximation algorithm on percentage of (i) valuable and (ii) highly observable islands under $B = 40$, $N = 5$ and $L = 20$.

	BC $\times$ power	BC	degree	power
(i)	54.55	48.82	48.30	47.07
(ii)	42.48	38.01	37.60	36.80

Configuration of proposed approximation algorithm: We compare performance of the proposed approximation algorithm under different limits on number (N) of paths in Fig. 5 and Table VI between each node and the CC. Figure 6 and Table VII show the effect of length (L) of paths between each node and the CC in the approximation algorithm. We also evaluate the approximation algorithm such that the paths in each P_i ($i \in V$) are disjoint with each other, without limitation on path length and number. The results note that topology estimation and cascade prevention achieve the best performance by allowing overlap between paths,

picking a smaller N and a larger L . For the rest of the results, we set $N = 5$ and $L = 20$ in the approximation algorithm. Also, Fig. 3(b) shows the designed communication network of the approximation algorithm for IEEE 118-bus system under these settings. In this figure, budget and node weight definition are considered 40 and “the BC of the node $\times$ the real power injected at the node”, respectively.

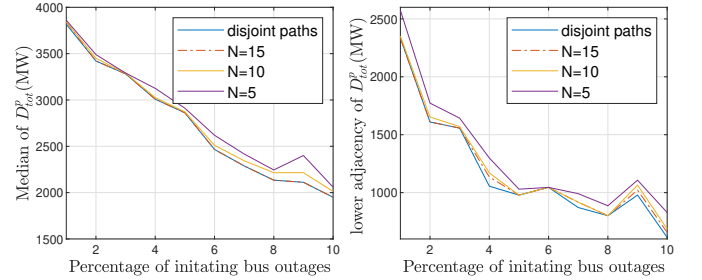


Figure 5: Performance evaluation of approximation algorithm in terms of median and lower adjacency of D_{tot}^P for different N under $B = 40$ and $L = 20$.

Table VI: The effect of varying N on percentage of (i) valuable and (ii) highly observable islands under $B = 40$ and $L = 20$.

	disjoint path	$N = 15$	$N = 10$	$N = 5$
(i)	42.83	48.82	51.79	54.78
(ii)	34.04	37.91	39.84	42.66

Table VII: The effect of varying L on percentage of (i) valuable and (ii) highly observable islands under $B = 40$ and $N = 5$.

	disjoint path	$L = 20$	$L = 15$	$L = 10$
(i)	45.85	58.65	47.41	46.05
(ii)	36.44	45.68	37.41	36.64

Comparison under different budget: In Fig. 7 and Table VIII, we show the proposed approximation algorithm’s performance under different budgets. As expected, the higher budget

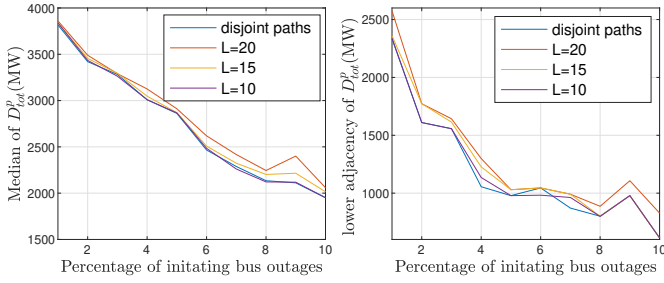


Figure 6: Performance evaluation of approximation algorithm in terms of median and lower adjacency of D_{tot}^p for different L under $B = 40$ and $N = 5$.

performs the best, as it produces a better-connected communication graph with more links, especially non-PLCC links.

To compare the proposed method with existing solutions where all the communication links are considered as non-PLCC, we compare the results from the motivating experiments (Section III-C) with Fig. 7 under 5% initial bus outages. The existing solutions require a budget of $B = 111.6$, and the mean value of the served power after cascade propagation is 3038.3 MW (70.81% of the initial power). Meanwhile, using the proposed technique, we can achieve a mean served power after cascade propagation of 2993.75 MW (69.78% of the initial power) at a budget of $B = 50$. This comparison shows that our solution can significantly reduce the cost of constructing the communication network with little negative impact on the efficacy of preventive control.

We also calculate the network's reliability as defined in (4a), under different budget scenarios. The results are 1.96, 0.84, 0.75 and 0.51 for budgets of 50, 45, 40 and 35, respectively. The results follow the same trend as Figs. 7 and note that a more reliable network delivers higher post-contingency served demand after cascade. The nodes' weights are normalized here.

To analyze the accuracy of the $\hat{B}$, Table IX provides the percentage of islands that have been correctly identified for different budget scenarios. For each scenario, 100 random cases were tested. In this table, the "slightly observable" islands are the valuable islands that are not highly observable; in other words, they include less than 50% nodes observable by the CC. As was expected, a higher budget leads to higher accuracy of island detection obtained from a more resilient communication network.

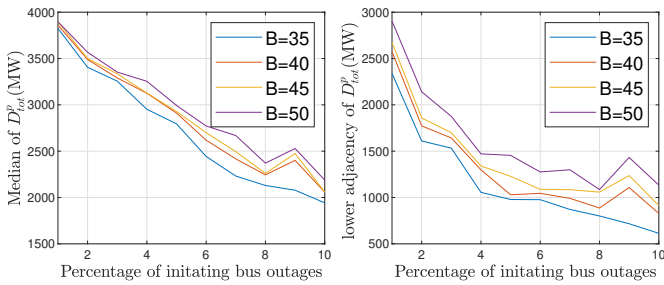


Figure 7: Performance evaluation of proposed approximation algorithm in terms of median and lower adjacency of D_{tot}^p for different B under $N = 5$ and $L = 20$.

Configuration of (α, β) -approximation of CMST algorithm: Recalling that the $(1 + \varepsilon, 1)$ -approximation of CMST algorithm

Table VIII: The effect of varying budget in proposed approximation algorithm on percentage of (i) valuable and (ii) highly observable islands under $N = 5$ and $L = 20$.

	$B = 35$	$B = 40$	$B = 45$	$B = 50$
(i)	32.97	43.40	52.74	64.80
(ii)	26.34	33.80	38.76	46.71

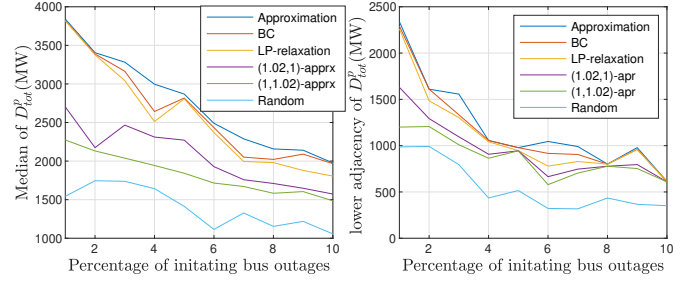


Figure 8: Performance evaluation of all algorithms in terms of median and lower adjacency of D_{tot}^p under $B = 40$, ($N = 5$ and $L = 20$ for approximation).

allows an extra budget of εB , hence, larger ε supplies a larger budget. We evaluated the communication network of $(1 + \varepsilon, 1)$ -approximation of CMST algorithm under varying ε , which are not shown in the paper. Results indicate that larger ε returns a graph with more links, especially more non-PLCC links; hence it performs better. To make the result of $(1 + \varepsilon, 1)$ -approximation of CMST algorithm comparable with other benchmarks, we fix ε to a small value of 0.02.

We also found that the $(1, 1 + \varepsilon)$ -approximation of CMST design is not sensitive to ε . The reason is that this method focuses on minimizing the cost of the constructed MST such that the total weight is at most B . In words, this algorithm groups links based on their cost, defined as "1 - reliability of the link". Then, in ascending order of the costs, it constructs a MST from the links' groups of not bigger than that particular cost. The cost of a non-PLCC link is smaller than PLCC; however, the MST of all non-PLCC links doesn't satisfy the weight budget constraint. So the $(1, 1 + \varepsilon)$ -approximation of CMST algorithm gives the same MST of PLCC links, independently of ε .

Figures 3(e/f) depicts the designed communication networks of $(1, 1.02)/(1.02, 1)$ -approximation of CMST algorithm for IEEE 118-bus system. Budget is 40 in these figures.

2) Overall Comparison of All Algorithms:

Finally, Fig. 8 and Table X compare the performance of all algorithms in terms of total post-contingency demand served D_{tot}^p and the percentage of valuable/highly observable islands in the power grid after cascade. Results show the proposed approximation algorithm consistently exceeds all the baselines, which emphasizes the importance of strategically placing links considering both the system topology and its

Table IX: The effect of varying budget in proposed approximation algorithm on percentage of estimation accuracy for (i) valuable, (ii) highly observable and (iii) slightly observable islands (under $N = 5$ and $L = 20$).

	$B = 35$	$B = 40$	$B = 45$	$B = 50$
(i)	87.48	90.87	95.99	99.57
(ii)	93.55	95.79	98.38	99.82
(iii)	52.20	67.23	87.32	98.57

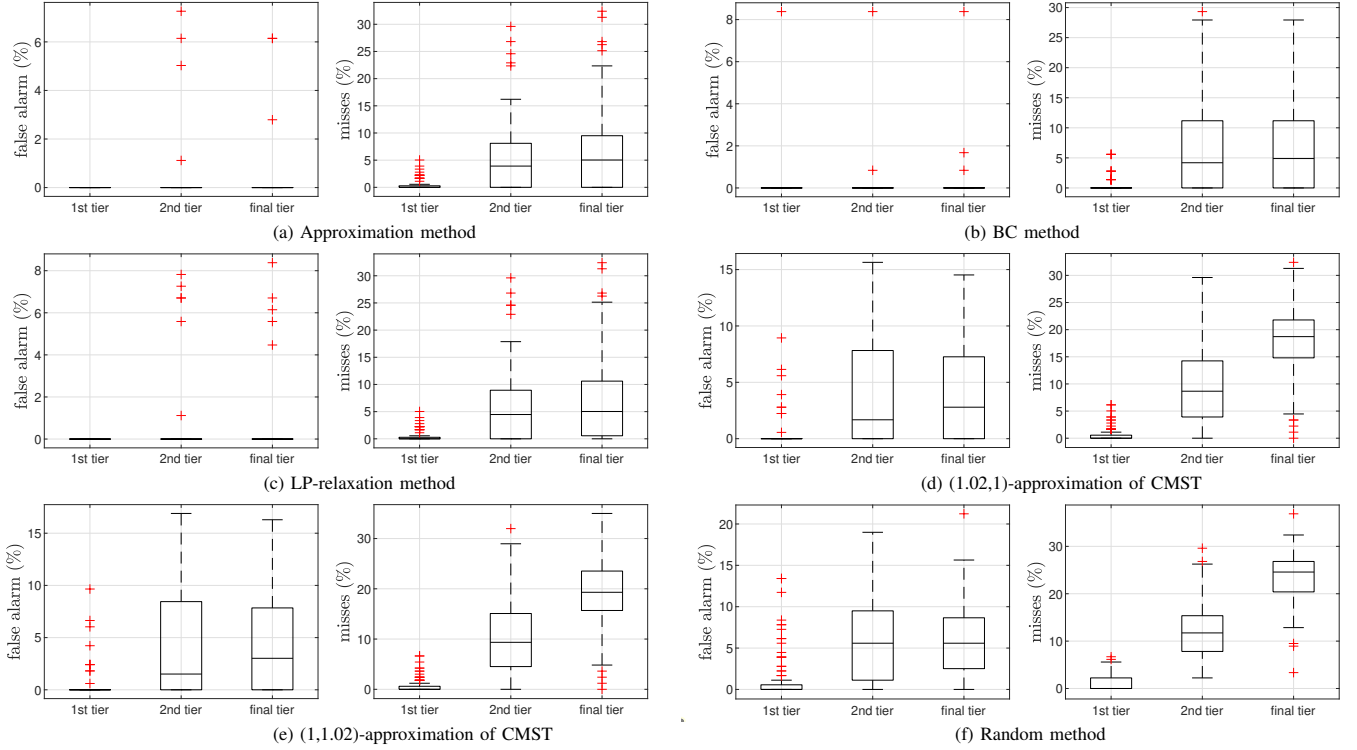


Figure 9: Performance evaluation of different methods in terms of percentage of misses and false alarms of line outage identification under $B = 40$, ($N = 5$ and $L = 20$ for proposed approximation method).

Table X: Performance of different methods in terms of (i) percentage of valuable and (ii) percentage of highly observable islands under $B = 40$, ($N = 5$ and $L = 20$ for approximation).

	Approximation	BC	LP-relaxation
(i)	43.46	38.95	39.15
(ii)	33.85	27.20	23.18
	(1.02,1)-aprx of CMST	(1,1.02)-aprx of CMST	Random
(i)	38.10	33.63	18.93
(ii)	15.39	10.23	3.05

Table XI: Performance evaluation of all algorithms in terms of percentage of estimation accuracy in (i) valuable, (ii) highly observable and (iii) slightly observable islands under $B = 40$, ($N = 5$ and $L = 20$ for approximation method).

	Approximation	BC	LP-relaxation
(i)	92.55	90.16	88.89
(ii)	96.22	94.23	93.62
(iii)	70.00	66.67	64.86
	(1.02,1)-aprx of CMST	(1,1.02)-aprx of CMST	Random
(i)	72.33	61.62	42.42
(ii)	89.62	74.29	60.40
(iii)	63.92	54.69	33.16

generation/load contribution. The second best algorithm, the BC method, outperforms the LP-relaxation with rounding benchmark and works notably better than (1, 1.02)/(1.02, 1)-approximation of CMST algorithms due to placing non-PLCC links at the network's weakest parts. In (1, 1.02)/(1.02, 1)-approximation of CMST designs, although there is not a notable gap in the percentage of valuable islands with respect to the LP-relaxation and BC methods, Only a small portion of these islands are highly observable. The reason is that these approximation of CMST algorithms create a spanning tree; hence the created graph is not well-connected and can be broken

into subtrees simply in failure occurrence. Furthermore, the (1, 1.02)-approximation of CMST graph only uses PLCC links that are not immune to failure, while (1.02, 1)-approximation of CMST picks non-PLCC links without any specific strategy.

Table XI gives the percentage of correctly identified islands for different methods. For each scenario, 100 random cases were tested, and 2% of initial node outages are considered. Results are in line with Table X which higher observability leads to higher accuracy of island identification.

Figure 9 illustrates the increase in the percentage of misses and false alarms as the cascade proceeds from one to two to the final tier of line outages. Misses are the cases in which certain lines are out but are identified as healthy. We compare these values for all benchmarks. For each scenario, 100 random cases and 2% of initial node outages are considered here. By comparing the first and final tier of line outages in all scenarios, it is logical to conclude that as the cascade progresses, the error accumulates, which further diminishes the accuracy of $\hat{B}$. Furthermore, the approximation method outperforms the other methods, especially against the random method, which highlights the importance of a well-designed communication network. Taking into account Figs. 8, 9 and Table X together, it seems that different methods follow the same trend in these two figures and the table. The origin of this stems from the resilience of the communication network against failures, both at the initiation and during the cascade. The more the communication network is observable by the CC, the higher the percentage of valuable/highly observable islands, and the more accurate $\hat{B}$ is. This, in turn, leads to better results in term of total post-contingency demand served

since the preventive controller produces a more accurate result in a closed-loop manner while solving (1), which depends on $\hat{\mathbf{B}}$.

VII. CONCLUSION & FUTURE WORK

We study the impact of coupling between the power grid and a SCADA-based communication network. Most modern SCADA systems use a variety of communication options within one system to meet their needs. Typically, there is no one-size-fits-all solution, so the SCADA system should be designed carefully to fit the system's needs. We combine the cost efficiency of PLCC links and the reliability of non-PLCC links and allocate a limited number of communication links while focusing on improving the robustness of such control system against cascading failures under the assumption of uncertain knowledge of failure and system topology. We not only proved the NP-hardness of the proposed solution in the general case, but we also developed a polynomial-time algorithm giving a constant approximation ratio under certain conditions. Extensive simulations on the 118-IEEE bus power system showed that the proposed algorithm achieves efficacy in estimating the system topology and different statistical measures of total demand served at the end of the cascade.

Finally, it is essential to remember that these technologies are not mutually exclusive. Different types of Non-PLCC and PLCC options can be used alone or in tandem, depending on the system's size and nature. Extension to more complex models is left to future work.

REFERENCES

- [1] "Supervisory control and data acquisition (scada) systems," *Technical information bulletin 04-1, National Communications Systems*, 2004. [Online]. Available: <https://www.cedengineering.com/userfiles/SCADA%20Systems.pdf>
- [2] X. Fan, S. G. Aksoy, Q. Huang, J. P. Ogle, D. Wang, A. Tbaileh, and T. Fu, "Coordination of transmission, distribution and communication systems for prompt power system recovery after disasters: report – grid and communication interdependency review and characterization of typical communication systems," 3 2019. [Online]. Available: <https://www.osti.gov/biblio/1526728>
- [3] H. Hadlach, H. Touijer, M. Zahri, M. El alami, and M. Habibi, "Modeling of a smart grid monitoring system using power line communication," in *2017 International Renewable and Sustainable Energy Conference (IRSEC)*, 2017, pp. 1–4.
- [4] K. Ali, A. X. Liu, I. Pefkianakis, and K. Kim, "Distributed spectrum sharing for enterprise powerline communication networks," in *2018 IEEE 26th International Conference on Network Protocols (ICNP)*, 2018, pp. 367–377.
- [5] A. Majumder and J. Caffery, "Power line communication: An overview," *Potentials, IEEE*, 2004.
- [6] I. H. Kim, A. Dabak, D. Rieken, and G. Gregg, "Evaluating the low-frequency power line communication in rural north america," *Texas Instruments*. [Online]. Available: www.ti.com/lit/wp/spr203/spr203.pdf.
- [7] S. Canale, A. Di Giorgio, A. Lanna, A. Mercurio, M. Panfil, and A. Pietrabissa, "Optimal planning and routing in medium voltage powerline communications networks," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 711–719, 2013.
- [8] S. Galli, A. Scaglione, and Z. Wang, "For the grid and through the grid: The role of power line communications in the smart grid," *Proceedings of the IEEE*, vol. 99, no. 6, pp. 998–1027, 2011.
- [9] H. Akkermans, D. Healey, and H. Ottosson, "The report on transmission of data over the electricity power lines." *AKMC & Spectrum & EnerSearch*, Europe, June 1998.
- [10] M. P. Sanders, "Power line carrier channel & application consideration for transmission line relaying," *Pulsar technologies*.
- [11] K. W. Louie, A. Wang, P. Wilson, and P. Buchanan, "Discussion on power line carrier applications," in *2006 Canadian Conference on Electrical and Computer Engineering*, 2006, pp. 655–658.
- [12] Q. Wang, M. Pipattanasomporn, M. Kuzlu, Y. Tang, Y. Li, and S. Rahman, "Framework for vulnerability assessment of communication systems for electric power grids," *IET Generation, Transmission Distribution*, vol. 10, no. 2, pp. 477–486, 2016.
- [13] "Power network telecommunication powerlink – power line carrier system." [Online]. Available: www.siemens.com
- [14] "Telecommunications for power utilities." [Online]. Available: <https://www.ge.com/digital/sites/default/files/download-assets/Utilities-Communications-Brochure-GE.pdf>
- [15] F. Aalamifar, "Viability of powerline communication for smart grid realization," 2012.
- [16] U. department of Energy, "Advanced metering infrastructure and customer systems," 2016. [Online]. Available: <https://www.energy.gov/sites/prod/files/2016/12/f34/AMI%20Summary%20Report-09-26-16.pdf>
- [17] M. Ouyang, "Review on modeling and simulation of interdependent critical infrastructure systems," *Reliability Engineering & System Safety*, vol. 121, pp. 43–60, 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0951832013002056>
- [18] J. W. Busby, K. Baker, M. D. Bazilian, A. Q. Gilbert, E. Grubert, V. Rai, J. D. Rhodes, S. Shidore, C. A. Smith, and M. E. Webber, "Cascading risks: Understanding the 2021 winter blackout in texas," *Energy Research Social Science*, vol. 77, p. 102106, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2214629621001997>
- [19] A. Berizzi, "The italian 2003 blackout," in *IEEE Power Engineering Society General Meeting, 2004.*, 2004, pp. 1673–1679 Vol.2.
- [20] V. Suresh, R. Sutradhar, S. Mandal, and K. Debnath, "Near miss of blackout in southern part of north eastern grid of india," in *2020 3rd International Conference on Energy, Power and Environment: Towards Clean Energy Technologies*, 2021, pp. 1–5.
- [21] Y. Cai, Y. Cao, Y. Li, T. Huang, and B. Zhou, "Cascading failure analysis considering interaction between power grids and communication networks," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 530–538, 2016.
- [22] M. Rahnamay-Naeini and M. M. Hayat, "Cascading failures in interdependent infrastructures: An interdependent markov-chain approach," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 1997–2006, 2016.
- [23] R. A. Shuvro, Z. Wangt, P. Das, M. R. Naeini, and M. M. Hayat, "Modeling cascading-failures in power grids including communication and human operator impacts," in *2017 IEEE Green Energy and Smart Systems Conference (IGESSC)*, 2017, pp. 1–6.
- [24] S. Mei, Y. Ni, G. Wang, and S. Wu, "A study of self-organized criticality of power system under cascading failures based on ac-opf with voltage stability margin," *IEEE Transactions on Power Systems*, vol. 23, no. 4, pp. 1719–1726, 2008.
- [25] M. H. Athari and Z. Wang, "Stochastic cascading failure model with uncertain generation using unscented transform," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 1067–1077, 2020.
- [26] J. Song, E. Cotilla-Sanchez, G. Ghanavati, and P. D. H. Hines, "Dynamic modeling of cascading failure in power systems," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2085–2095, 2016.
- [27] Y. Dai, M. Noebels, M. Panteli, and R. Preece, "Comparing dynamic and static modelling of cascading failures in power systems," in *The 12th Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion (MEDPOWER 2020)*, vol. 2020, 2020, pp. 301–307.
- [28] B. Carreras, V. Lynch, I. Dobson, and D. Newman, "Critical points and transitions in an electric power transmission model for cascading failure blackouts," *Chaos (Woodbury, N.Y.)*, vol. 12, pp. 985–994, 01 2003.
- [29] J. Yan, Y. Tang, H. He, and Y. Sun, "Cascading failure analysis with dc power flow model and transient stability analysis," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 285–297, 2015.
- [30] M. Korkali, J. Veneman, B. Tivnan, J. Bagrow, and P. Hines, "Erratum: Reducing cascading failure risk by increasing infrastructure network interdependence," *Scientific Reports*, vol. 8, p. 46959, 03 2018.
- [31] D. Z. Tootaghaj, N. Bartolini, H. Khamfroush, T. He, N. R. Chaudhuri, and T. L. Porta, "Mitigation and recovery from cascading failures in interdependent networks under uncertainty," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 2, pp. 501–514, 2019.
- [32] V. Farhadi, S. G. Vennelaganti, T. He, N. R. Chaudhuri, and T. L. Porta, "Budget-constrained reinforcement of scada for cascade mitigation," in *2021 International Conference on Computer Communications and Networks (ICCCN)*, 2021, pp. 1–9.

- [33] P. Musil, P. Mlýnek, J. Slacik, and J. Pokorný, "Simulation-based evaluation of the performance of broadband over power lines with multiple repeaters in linear topology of distribution substations," *Applied Sciences*, vol. 10, p. 6879, 2020.
- [34] H. Ferreira, I. HC Ferreira, L. Lampe, JE Newbury and TG Swart, Editors: "Power Line Communications," John Wiley and Sons, UK, ISBN 978-0-470-740309, June 2010., 01 2010.
- [35] W. C. Black, "Data transmission through distribution transformers without bypass components," in *ISPLC2010*, 2010, pp. 13–17.
- [36] J. Ushida, M. Tokushima, M. Shirane, A. Gomyo, and H. Yamada, "Immittance matching for multidimensional open-system photonic crystals," vol. 68, no. 15, p. 155115, Oct. 2003.
- [37] R. Bixby, "A brief history of linear and mixed-integer programming computation," *Documenta Mathematica*, 01 2012.
- [39] J. Kronqvist and A. Lundell, "Convex minlp – an efficient tool for design and optimization tasks?" *Computer Aided Chemical Engineering*, 2019.
- [40] C. Ruben, S. C. Dhulipala, A. Suman Bretas, and N. Bretas, "Optimal pmu allocation on smart grids: a milp model considering minimal current measurements," in *2018 IEEE Power Energy Society General Meeting (PESGM)*, 2018, pp. 1–5.
- [41] M. Parandehgheibi, E. Modiano, and D. Hay, "Mitigating cascading failures in interdependent power grids and communication networks," in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2014, pp. 242–247.
- [42] W. H. Al-Rousan, C. Wang, and F. Lin, "A discrete event theory based approach for modeling power system cascading failures," in *2019 IEEE Power Energy Society General Meeting (PESGM)*, 2019, pp. 1–5.
- [43] P. Hines, I. Dobson, and P. Rezaei, "Cascading power outages propagate locally in an influence graph that is not the actual grid topology," in *2017 IEEE Power Energy Society General Meeting*, 2017, pp. 1–1.
- [44] V. S. Rajkumar, M. Tealane, A. Ştefanov, A. Presekal, and P. Palensky, "Cyber attacks on power system automation and protection and impact analysis," in *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, 2020, pp. 247–254.
- [45] M. Ni and M. Li, "Reliability assessment of cyber physical power system considering communication failure in monitoring function," in *2018 International Conference on Power System Technology (POWERCON)*, 2018, pp. 3010–3015.
- [46] Z. Dong, M. Tian, and J. Liang, "Cascading failures of spatially embedded cyber physical power system under localized attacks," in *2018 37th Chinese Control Conference (CCC)*, 2018, pp. 6154–6159.
- [47] S. G. Vennelaganti, N. R. Chaudhuri, T. He, and T. L. Porta, "Topology estimation following islanding and its impact on preventive control of cascading failure," 2021, arXiv:2104.06473.
- [48] [Online]. Available: <https://icseg.iti.illinois.edu/ieee-118-bus-system/>
- [49] C. Report, "Requirements and performance of packet switching networks with special reference to telecontrol," August (1991). [Online]. Available: <https://cigreindia.org/CIGRE%20Lib/Tech.%20Brochure/>
- [50] L. Di Puglia Pugliese, M. Gaudioso, F. Guerriero, and G. Miglionico, "An algorithm to find the link constrained steiner tree in undirected graphs," in *Mathematical Software – ICMS 2016*, G.-M. Greuel, T. Koch, P. Paule, and A. Sommese, Eds. Cham: Springer International Publishing, 2016, pp. 492–497.
- [51] Y. Chen, "Polynomial time approximation schemes for the constrained minimum spanning tree problem," *Journal of Applied Mathematics*, vol. 2012, Special Issue, 03 2012.
- [52] R. Ravi and M. Goemans, "The constrained minimum spanning tree problem (extended abstract)," in *SWAT*, 1996.
- [53] J. Golbeck, "Chapter 21 - analyzing networks," in *Introduction to Social Media Investigation*, J. Golbeck, Ed. Boston: Syngress, 2015, pp. 221 – 235. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/B9780128016565000214>
- [54] E. W. Weisstein, "Norm." *From MathWorld—A Wolfram Web Resource*. [Online]. Available: <https://mathworld.wolfram.com/Norm.html>
- [55] D. J. Marihart, "Communications technology guidelines for ems/scada systems," *IEEE Transactions on Power Delivery*, vol. 16, no. 2, pp. 181–188, 2001.
- [56] A. E. Jahromi and Z. B. Rad, "Optimal topological design of power communication networks using genetic algorithm," *Scientia Iranica*, vol. 20, pp. 945–957, 2013.
- [57] S. Ioannidis and E. Yeh, "Jointly optimal routing and caching for arbitrary network topologies," 2017, arXiv:1708.05999.

Proof.

Definition A.1 ([57]). *Goemans-Williamson inequality gives the following bound (11) for any sequence of $y_i \in [0, 1]$, $i \in \{1, \dots, n\}$:*

$$\left(1 - \frac{1}{e}\right) \min\left\{1, \sum_{i=1}^n y_i\right\} \leq 1 - \prod_{i=1}^n (1 - y_i) \leq \min\left\{1, \sum_{i=1}^n y_i\right\} \quad (11)$$

By applying (11) to the first term in (6), for all nodes in V_1 we have

$$\begin{aligned} \sum_{i \in V_1} \gamma_i \left(1 - \frac{1}{e}\right) \min\left\{1, \sum_{m \in P_i} \prod_{l \in m} A_l\right\} &\leq \sum_{i \in V_1} \gamma_i \\ \left(1 - \prod_{m \in P_i} (1 - \prod_{l \in m} A_l)\right) &\leq \sum_{i \in V_1} \gamma_i \min\left\{1, \sum_{m \in P_i} \prod_{l \in m} A_l\right\}. \end{aligned} \quad (12)$$

As for $i \in V_1$: $\sum_{m \in P_i} \prod_{l \in m} A_l \leq \sum_{m \in P_i} \prod_{l \in m} p_{max} \leq 1$, so $\min\left\{1, \sum_{m \in P_i} \prod_{l \in m} A_l\right\} = \sum_{m \in P_i} \prod_{l \in m} A_l$. Thus,

$$\begin{aligned} \left(1 - \frac{1}{e}\right) \sum_{i \in V_1} \gamma_i \sum_{m \in P_i} \prod_{l \in m} A_l &\leq \sum_{i \in V_1} \gamma_i \times \\ \left(1 - \prod_{m \in P_i} (1 - \prod_{l \in m} A_l)\right) &\leq \sum_{i \in V_1} \gamma_i \sum_{m \in P_i} \prod_{l \in m} A_l. \end{aligned} \quad (13)$$

Let $|m_{1,max}|$ denotes the maximum hop count per path over all the paths in $\bigcup_{i \in V_1} P_i$. Define

$$\epsilon := (e - (e - 1)(1 - p_{min}))^{|m_{1,max}| - 1}. \quad (14)$$

If $(1 - p_{min})|m_{1,max}| < 1$, then $\sum_{l \in m} (1 - A_l) < 1$ for all $m \in \bigcup_{i \in V_1} P_i$, $\epsilon \in [\frac{1}{e}, 1)$, and

$$\epsilon \geq \left(e - (e - 1) \sum_{l \in m} (1 - A_l)\right)^{-1}, \quad \forall m \in \bigcup_{i \in V_1} P_i. \quad (15)$$

By applying (11) to (13), we have

$$\begin{aligned} LB &:= \left(1 - \frac{1}{e}\right) \sum_{i \in V_1} \gamma_i \sum_{m \in P_i} \left(1 - \sum_{l \in m} (1 - A_l)\right) \leq \\ &\sum_{i \in V_1} \gamma_i \left(1 - \prod_{m \in P_i} (1 - \prod_{l \in m} A_l)\right) \leq \\ &\sum_{i \in V_1} \gamma_i \sum_{m \in P_i} \left(\left(1 - (1 - \frac{1}{e}) \sum_{l \in m} (1 - A_l)\right)\right) := UB. \end{aligned} \quad (16)$$

From (15), we get

$$\begin{aligned} &\left(1 - \frac{1}{e}\right) \left(1 - \sum_{l \in m} (1 - A_l)\right) \\ &\geq (1 - \epsilon) \left(1 - (1 - \frac{1}{e}) \sum_{l \in m} (1 - A_l)\right), \end{aligned} \quad (17)$$

APPENDIX

PROOF OF THEOREM 2:

which implies

$$LB \geq (1 - \epsilon) \sum_{i \in V_1} \gamma_i \sum_{m \in P_i} \left(1 - \left(1 - \frac{1}{e} \right) \sum_{l \in m} (1 - A_l) \right) = (1 - \epsilon) UB. \quad (18)$$

Hence,

$$(1 - \epsilon) UB \leq \sum_{i \in V_1} \gamma_i \left(1 - \prod_{m \in P_i} (1 - \prod_{l \in m} A_l) \right) \leq UB. \quad (19)$$

By applying (11) to the second term in (6), for all nodes in V_2 we have

$$\begin{aligned} \sum_{i \in V_2} \gamma_i \left(1 - \frac{1}{e} \right) \min\{1, \sum_{m \in P_i} \prod_{l \in m} A_l\} &\leq \sum_{i \in V_2} \gamma_i \times \\ \left(1 - \prod_{m \in P_i} (1 - \prod_{l \in m} A_l) \right) &\leq \sum_{i \in V_2} \gamma_i \min\{1, \sum_{m \in P_i} \prod_{l \in m} A_l\}. \end{aligned} \quad (20)$$

As for all nodes in V_2 , $\sum_{m \in P_i} \prod_{l \in m} A_l \geq \prod_{l \in m_{2, \max}} p_{\min} \geq 1$, so $\min\{1, \sum_{m \in P_i} \prod_{l \in m} A_l\} = 1$. Thus,

$$\begin{aligned} \left(1 - \frac{1}{e} \right) \sum_{i \in V_2} \gamma_i &\leq \sum_{i \in V_2} \gamma_i \left(1 - \prod_{m \in P_i} (1 - \prod_{l \in m} A_l) \right) \leq \\ \sum_{i \in V_2} \gamma_i &:= UB'. \end{aligned} \quad (21)$$

From (15), $\frac{1}{e} \leq \epsilon$. Hence,

$$(1 - \epsilon) UB' \leq \sum_{i \in V_2} \gamma_i \left(1 - \prod_{m \in P_i} (1 - \prod_{l \in m} A_l) \right) \leq UB'. \quad (22)$$

Considering the third term in (6), for all nodes in $V \setminus \{V_1 \cup V_2\}$ we have (23), where $|m_{3, \max}|$ denotes the maximum hop count per path over all the paths in $\bigcup_{i \in V \setminus \{V_1 \cup V_2\}} P_i$.

$$\begin{aligned} LB'' &:= \sum_{i \in V \setminus \{V_1 \cup V_2\}} \gamma_i \prod_{l \in m_{3, \max}} p_{\min} \leq \\ \sum_{i \in V \setminus \{V_1 \cup V_2\}} \gamma_i &\left(1 - \prod_{m \in P_i} (1 - \prod_{l \in m} A_l) \right) \leq \\ \sum_{i \in V \setminus \{V_1 \cup V_2\}} \gamma_i &\left(1 - \prod_{m \in P_i} (1 - \prod_{l \in m} p_{\max}) \right) := UB'' \end{aligned} \quad (23)$$

Special case: Consider the condition of $p_{\min}^{|m_{3, \max}|} \geq p_{\max}^{|m_{3, \min}|} (1 - \frac{1}{e}) |P_{\max}|$, where $|m_{3, \min}|$ defines the minimum hop count per path over all the paths in $\bigcup_{i \in V \setminus \{V_1 \cup V_2\}} P_i$ and $|P_{\max}| := \max_{i \in V \setminus \{V_1 \cup V_2\}} |P_i|$.

In this special case, by applying binomial expansion, we get

$$\begin{aligned} UB'' &= \sum_{i \in V \setminus \{V_1 \cup V_2\}} \gamma_i \left(1 - \prod_{m \in P_i} (1 - \prod_{l \in m} p_{\max}) \right) \\ &\leq \sum_{i \in V \setminus \{V_1 \cup V_2\}} \gamma_i \left(1 - \left(1 - \prod_{l \in m_{3, \min}} p_{\max} \right)^{|P_i|} \right). \end{aligned} \quad (24)$$

So,

$$\begin{aligned} \sum_{i \in V \setminus \{V_1 \cup V_2\}} \gamma_i &\left(1 - \left(1 - \prod_{l \in m_{3, \min}} p_{\max} \right)^{|P_i|} \right) < \\ \sum_{i \in V \setminus \{V_1 \cup V_2\}} \gamma_i &\left(1 - \left(1 - |P_i| \prod_{l \in m_{3, \min}} p_{\max} \right) \right) = \\ \sum_{i \in V \setminus \{V_1 \cup V_2\}} \gamma_i &|P_i| p_{\max}^{|m_{3, \min}|} \leq \left(1 - \frac{1}{e} \right)^{-1} \times \\ \sum_{i \in V \setminus \{V_1 \cup V_2\}} \gamma_i &p_{\min}^{|m_{3, \max}|} = \left(1 - \frac{1}{e} \right)^{-1} LB''. \end{aligned} \quad (25)$$

From (15), $\frac{1}{e} \leq \epsilon$. Hence,

$$\begin{aligned} (1 - \epsilon) UB'' &\leq \\ \sum_{i \in V \setminus \{V_1 \cup V_2\}} \gamma_i &\left(1 - \prod_{m \in P_i} (1 - \prod_{l \in m} A_l) \right) \leq UB''. \end{aligned} \quad (26)$$

Finally, by (19), (22) and (26), we have

$$(1 - \epsilon)(UB + UB' + UB'') \leq (4a) \leq UB + UB' + UB''. \quad (27)$$

□



Vajiheh Farhadi received the M.S. degree in electrical engineering in Iran, 2013 and the Ph.D. degree in computer science and engineering from Pennsylvania State University, PA, in 2022, advised by Prof. Thomas La Porta and Prof. Ting He. Her research interests include resource allocation, optimization algorithms, cellular network, control network of the smart grid, interdependent power grid SCADA system.



Nilanjan Ray Chaudhuri (S'08-M'09-SM'16) received the Ph.D. degree in power systems from Imperial College London, London, UK in 2011. From 2005 to 2007, he worked in General Electric (GE) John F. Welch Technology Center. He came back to GE and worked in GE Global Research Center, NY, USA as a Lead Engineer during 2011 to 2014. Presently, he is an Associate Professor with the School of Electrical Engineering and Computer Science at Penn State, University Park, PA. He was an Assistant Professor with North Dakota State University, Fargo, ND, USA during 2014-2016. He is a member of the IEEE and IEEE PES. Dr. Ray Chaudhuri is the lead author of the book *Multi-terminal Direct Current Grids: Modeling, Analysis, and Control* (Wiley/IEEE Press, 2014). He served as an Associate Editor of the IEEE TRANSACTIONS ON POWER DELIVERY (2013 – 2019) and IEEE PES LETTERS (2016 - present). Dr. Ray Chaudhuri was the recipient of the National Science Foundation Early Faculty CAREER Award in 2016 and Joel and Ruth Spira Excellence in Teaching Award in 2019.



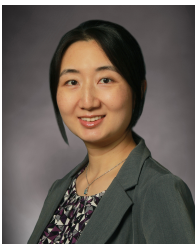
Sai Gopal Vennelaganti (Member, IEEE) received the bachelor's degree in electrical engineering from the Indian Institute of Technology Madras, Chennai, India, in 2016, and the Ph.D. degree from Pennsylvania State University, Pennsylvania, PA, USA, in 2020. He is currently an Engineer or Scientist with Grid Operations and Planning Research and Development Group, EPRI, Knoxville, TN, USA. His research interests include modeling and prevention of cascading failure in cyber-physical power systems, topology estimation, power system stability, and

control applications in HVDC, multiterminal HVDC, power quality, hybrid microgrids, and smart grids.



Thomas F. La Porta is the Director of the School of Electrical Engineering and Computer Science and Penn State University. He is an Evan Pugh Professor and the William E. Leonhard Chair Professor in the Computer Science and Engineering Department and the Electrical Engineering Department. He received his B.S.E.E. and M.S.E.E. degrees from The Cooper Union, New York, NY, and his Ph.D. degree in Electrical Engineering from Columbia University, New York, NY. He joined Penn State in 2002. He was the founding Director of the Institute of Networking

and Security Research at Penn State. Prior to joining Penn State, Dr. La Porta was with Bell Laboratories for 17 years. He was the Director of the Mobile Networking Research Department in Bell Laboratories, Lucent Technologies where he led various projects in wireless and mobile networking. He is an IEEE Fellow, Bell Labs Fellow, and received the Bell Labs Distinguished Technical Staff Award. He also won two Thomas Alva Edison Patent Awards. Dr. La Porta was the founding Editor-in-Chief of the IEEE Transactions on Mobile Computing. He served as Editor-in-Chief of IEEE Personal Communications Magazine. He was the Director of Magazines for the IEEE Communications Society and was on its Board of Governors for three years.



Ting He (SM'13) received the B.S. degree in computer science from Peking University, China, in 2003 and the Ph.D. degree in electrical and computer engineering from Cornell University, Ithaca, NY, in 2007. Dr. He is an Associate Professor in the School of Electrical Engineering and Computer Science at the Pennsylvania State University, University Park, PA. Her interests include computer networking, performance evaluation, statistical inference, and machine learning.

Dr. He is a senior member of IEEE. She is an Associate Editor for IEEE Transactions on Communications (2017-2020) and IEEE/ACM Transactions on Networking (2017-2023). She has served as the TPC Co-Chair of IEEE ICCN (2022), an Area TPC Chair of IEEE INFOCOM (2021), and a TPC member of many international conferences in the areas of networking and distributed computing. She was the Membership Co-Chair of ACM N2Women in 2013-2014 and was listed in "N2Women: Rising Stars in Networking and Communications" in 2017. Her work received the Research Division Award and multiple Outstanding Contributor Awards from IBM, the Military Impact and the Commercial Prosperity Awards from DAIS-ITA, the Most Collaboratively Complete Publications Award from NIS-ITA, the 2021 IEEE Communications Society Leonard G. Abraham Prize, the Best Paper Award at the 2013 International Conference on Distributed Computing Systems (ICDCS), the Outstanding Student Paper Award at the 2015 ACM SIGMETRICS, and the Best Student Paper Award at the 2005 International Conference on Acoustic, Speech and Signal Processing (ICASSP).