Localization & Mitigation of Cascading Failures in Power Systems, Part III: Real-time Mitigation

Linqi Guo, Chen Liang, Alessandro Zocca, Steven H. Low, and Adam Wierman

***Abstract*— Cascading failures in power systems propagate non-locally, making the control of outages extremely difficult. In Part III of this work, we leverage the properties of tree partitioning developed in Parts I and II to propose a distributed control strategy that offers strong guarantees in both the mitigation and localization of cascading failures. Specifically we adopt a recently developed distributed frequency regulation approach, called the Unified Control, that integrates primary and secondary control as well as congestion management at frequency control timescale. When the balancing areas over which the Unified Control operates form a tree partition, our proposed strategy will regulate the system to a steady state where the impact of initial line outages is localized within the areas where they occur whenever possible and stop the cascading process. When initial line outages cannot be localized, the proposed strategy provides a configurable design that involves and coordinates progressively more balancing areas for failure mitigation in a way that can be optimized for different priorities. We compare the proposed control strategy with the classical automatic generation control (AGC) on the IEEE 118-bus and 2736-bus test networks. Simulation results show that our strategy greatly improves overall reliability in terms of the $N-k$ security standard, and localizes the impact of initial failures in majority of the load profiles that are examined. Moreover, the proposed framework incurs significantly less load loss, if any, compared to AGC, in all of our case studies.**

I. Introduction

Grid reliability is traditionally enhanced through redundant transmission lines that can maintain connectivity in the face of line outages. Higher connectivity however can also exacerbate the propagation of component failures, potentially leading to large-scale blackouts. In Parts I [1] and II [2] of this paper, we propose a complementary approach to grid reliability that switch off certain existing lines to partition the grid into regions that are connected in a tree topology. Tree partitioning will contain the impact of line failures within each region, thus reducing the risk of large-scale blackouts.

Parts I and II focus on establishing failure localization as a general property of tree partitioning independent of generation and load injections. We therefore adopt a perspective that assumes injections keep unchanged after a non-cut failure (and are changed only after a bridge failure according to a generic balancing rule that re-balances power in each island). Although the tree-partition properties can be of independent interest for various applications, this perspective is both unrealistic and too pessimistic for failure mitigation. It does not leverage the fact that frequency control mechanisms adjust injections of controllable generators and loads immediately in response to a line outage, at a faster timescale than post-contingency line tripping (usually from thermal processes). In Part III, we supplement the cascading failure model of Parts I and II to capture the mechanism through which fast timescale dynamics affects power flow redistribution and failure propagation. The integrated failure model is not only more realistic, but also offers an additional means to mitigate cascading failure through a better design of the frequency control mechanism.

Contributions of Part III of this paper: *We integrate a distributed frequency control strategy with tree partitioning to provide provable failure mitigation and localization guarantees.* To the best of our knowledge, this is the first attempt to leverage results from the frequency regulation literature in the context of cascading failure, bringing new perspectives and insights to both literatures. Our control strategy guarantees that, whenever it is feasible to avoid it, failures do not propagate (see Section II for a rigorous definition), and that the impact of failures are localized as much as possible in a manner configurable by the system operator.

We introduce the main idea of this new control strategy in Section III. The key piece of our design builds upon the Unified Controller (UC), a recent mechanism developed in the frequency regulation literature [3]–[7]. We specifically leverage the ability of UC to enforce line limits at a fast timescale whenever possible. Our design revolves around the properties that emerge when the balancing areas that UC manages form a tree partition. More specifically, in Section IV, we characterize how UC responds to an initial failure when it operates over a tree partition, and prove that a non-critical failure is automatically mitigated and localized. Later, in Section V, we discuss how tree partitioning enables the system operator to configure its mitigation strategy to minimize the impact of critical failures, and prove that UC can be extended to detect such scenarios as part of its normal operation.

In order to establish these results, we propose an integrated failure propagation model in Section II, which lies at the interface between fast timescale frequency dynamics and slow timescale line tripping process. Further, we prove new results on the UC optimization problem using the spectral representation of DC power flow equations established in Part I of this work [1]. Lastly, we apply classical results from convex optimization to show that critical failures can always be detected in a distributed fashion.

In Section VI, we compare the proposed control strategy (tree partitioning + UC) with classical Automatic Generation Control (AGC) using the IEEE 118-bus and 2736-bus test networks. We demonstrate that by switching off only a

This work has been supported by Resnick Fellowship, Linde Institute Research Award, NWO Rubicon grant 680.50.1529, NSF through grants CCF 1637598, ECCS 1619352, ECCS 1931662, CNS 1545096, CNS 1518941, CPS ECCS 1739355, CPS 154471.

LG, CL, SHL, AW are with the Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, 91125, USA. Email: {lguo, cliang2, slow, adamw}@caltech.edu. AZ is with the Department of Mathematics of the Vrije Universiteit Amsterdam, 1081HV, The Netherlands. Email: a.zocca@vu.nl.

small number of transmission lines and adopting UC as the fast timescale controller, one can significantly improve overall grid reliability in terms of the $N-k$ security standard. Moreover, in a majority of the load profiles that are examined, our control strategy localizes the impact of initial failures to the regions where they occur, leaving the operating points of all other balancing areas unchanged. This decoupling property across balancing areas can be important in practice. Lastly, we highlight that when load shedding is necessary, the proposed strategy incurs significantly smaller load loss.

Load shedding to mitigate cascading failure has been studied in the literature, e.g., using reinforcement learning [8], by shedding only the interruptible part of each load [9], adaptively using affine control based on observed states [10], formulated as a DC OPF problem [11], or by treating the cascading process as a discrete-time optimal control problem [12]. DC power flow model is used in [10]–[12] as in this paper. Unlike these works, however, we do not propose separate load shedding schemes to be implemented at the slow (power flow) timescale. Instead, load and generation control under the UC framework is performed as part of frequency regulation in fast timescale, during both normal operation and in a contingency.

II. Integrated Failure Propagation Model

The failure propagation model in Part II assumes the injections remain unchanged after a non-cut failure (and change only after a bridge failure). This is both unrealistic and too pessimistic since this perspective ignores fast timescale dynamics that respond to line failures immediately, which can potentially drive the system to a post-contingency equilibrium with different injections.

To account for this, we present an integrated cascading failure model that incorporates frequency control dynamics and generalizes the steady-state DC failure model in Part II. This model offers more flexibility in controller design and enables us to address challenges in failure mitigation using tools from the frequency regulation literature.

A. Fast Timescale Dynamics

Consider a power transmission network described by the graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$. Using the definitions in Table I where most variables represent deviations from their pre-contingency nominal values, we describe the post-contingency linearized frequency dynamics by the standard model:

$$\dot{\theta}_j = \omega_j, \qquad j \in \mathcal{N} \quad \text{(1a)}$$

$$M_j\dot{\omega}_j = r_j + d_j - D_j\omega_j - \sum_{e\in\mathcal{E}} C_{je} f_e, \quad j \in \mathcal{N} \quad \text{(1b)}$$

$$f_{ij} = B_{ij}(\theta_i - \theta_j), \qquad (i,j) \in \mathcal{E}. \quad \text{(1c)}$$

The above differential-algebraic equations model the fast timescale response of the system to a transmission line failure. The post-contingency injection deviation $p_j(t) := r_j + d_j(t)$ consists of the post-contingency disruption r_j and system response $d_j(t)$. The vector $d(t) := (d_j(t), j \in \mathcal{N})$ models frequency control and their values are determined by a feedback control law (an uncontrollable constant-power load is simply a special case where the control law sets corresponding $d_j(t) \equiv d_j$). We assume in this paper that the feedback controller is stabilizing and drives the closed-loop system towards an equilibrium as long as the post-contingency disruption $r := (r_j, j \in \mathcal{N})$ can be feasibly mitigated (see Section V for more discussion).

TABLE I: Variables associated with buses and transmission lines.

$\theta := (\theta_j, j \in \mathcal{N})$	bus voltage angle deviations from pre-contingency values
$\omega := (\omega_j, j \in \mathcal{N})$	bus frequency deviations from pre-contingency values
$r := (r_j, j \in \mathcal{N})$	system disturbances
$d := (d_j, j \in \mathcal{N})$	power injection deviations from pre-contingency values for generator buses; controllable load deviations from pre-contingency values for load buses
$p := (p_j, j \in \mathcal{N})$	aggregate post-contingency injection deviation
$\overline{d}_j, \underline{d}_j, j \in \mathcal{N}$	upper and lower limits for the adjustable injection d_j
$D_j\omega_j, j \in \mathcal{N}$	aggregate generator damping for generator buses; aggregate load frequency response for load buses
$M_j, j \in \mathcal{N}$	inertia constants
$f := (f_e, e \in \mathcal{E})$	branch flow deviations from pre-contingency values
$\overline{f}_e, \underline{f}_e, e \in \mathcal{E}$	upper and lower limits for branch flow deviations
$n := \lvert\mathcal{N}\rvert$	number of buses
$m := \lvert\mathcal{E}\rvert$	number of transmission lines
$C \in \mathbb{R}^{n\times m}$	post-contingency incidence matrix of $\mathcal{G}$: $C_{je} = 1$ if j is the source of e, $C_{je} = -1$ if j is the destination of e, and $C_{je} = 0$ otherwise
$B := \mathrm{diag}(B_e, e \in \mathcal{E})$	branch flow linearization coefficients that depend on line susceptances, nominal voltage magnitudes and reference phase angles

Definition 1. *A state $x^* := (\theta^*, \omega^*, d^*, f^*) \in \mathbb{R}^{3n+m}$ is said to be a* **closed-loop equilibrium** *or simply an* **equilibrium** *of* (1) *if the right hand sides of* (1a)(1b) *are zero and* (1c) *is satisfied at x^*.

The frequency dynamics (1) implies that an equilibrium x^* satisfies

$$w^* = 0, \quad p^* = r + d^* = Cf^*, \quad f^* = BC^T\theta^*.$$

In other words, x^* is a solution of the DC power flow model.[1]

The equilibrium to which the closed-loop system (1) converges determines the post-contingency DC power flow solution and therefore the cascading process. By explicitly modeling fast timescale frequency control dynamics as part of the cascading process, our model offers more flexibility in controller design. Different choices of $d(t)$ induce different cascading failure models. For instance, as shown in Appendix VI, if we adopt droop control for $d(t)$, the failure model in Part II of the paper (where injections do not change after a non-cut failure, and a bridge failure impact the injections following a certain power balancing rule $\mathcal{R}$) can be readily recovered. As another example, if AGC is adopted for $d(t)$, the cascading process will unfold in a way where injections and line flows are changed even after a

[1] In primary frequency control literature (see [3], [5] for instance), the right hand side of (1a) is not required to be zero for an equilibrium point x^*. We impose this requirement on (1a) here as our discussion focuses on controllers that achieve secondary frequency control and thus $\omega^* = 0$ always holds. Our model and results can be readily extended to the case where $\omega^* \neq 0$; see Appendix VI for more details.

non-cut failure. Since traditional AGC does not enforce line limits (congestion is managed at a slow timescale), some lines may carry flows above their thermal limits and are tripped subsequently.

This integrated model offers an additional means to mitigate cascading failures through a better design of the frequency control mechanism. Our proposed approach leverages this extra freedom and adopts a recent frequency control approach known as Unified Controller (UC) for $d(t)$. In contrast to traditional AGC, UC drives the closed-loop system to an equilibrium that respects line limits whenever possible. We will show in Section IV that it will in fact localize the impact even of bridge failures.

A key insight of [3]–[7] is that a closed-loop equilibrium x^* of (1) is also an optimal solution of a certain optimization problem that can be determined explicitly. Different frequency controllers $d(t)$ induce different dynamics (1), whose closed-loop equilibria solve optimization problems with corresponding objective functions and constraints. As such, different frequency controllers can alternatively be modelled by the underlying optimization problems that their equilibria solve. We take this perspective when we discuss the design of UC in the following sections.

B. Unified Controller (UC)

UC is a control approach recently proposed in the frequency regulation literature [3]–[7]. Compared to classical droop control or Automatic Generation Control (AGC) [13], UC integrates primary control, secondary control, and congestion management simultaneously at the frequency control timescale. The key feature of UC we use here is that the *closed-loop* equilibrium of (1) under UC solves the following optimization on the post-contingency network:

$$\min_{\theta,\omega,d,f} \quad \sum_{j\in\mathcal{N}} c_j(d_j) \tag{2a}$$

$$\text{s.t.} \quad \omega = 0, \tag{2b}$$

$$r + d - Cf = 0, \tag{2c}$$

$$f = BC^T\theta, \tag{2d}$$

$$ECf = 0, \tag{2e}$$

$$\underline{f}_e \le f_e \le \overline{f}_e, \quad e \in \mathcal{E}, \tag{2f}$$

$$\underline{d}_j \le d_j \le \overline{d}_j, \quad j \in \mathcal{N}, \tag{2g}$$

where $c_j(\cdot)$'s are associated cost functions that penalize deviations from the last optimal dispatch (and hence attain minimum at $d_j = 0$), (2b) ensures secondary frequency regulation is achieved, (2c) guarantees power balance at each bus, (2d) is the DC power flow equation, (2e) enforces zero area control error [13], (2f) and (2g) are the flow and control limits. The matrix E encodes balancing area information as follows. Given a partition $\mathcal{P}^{\text{UC}} = \{\mathcal{N}_1, \mathcal{N}_2, \cdots, \mathcal{N}_k\}$ of $\mathcal{G}$ that specifies the balancing areas in secondary frequency control, $E \in \{0,1\}^{|\mathcal{P}^{\text{UC}}|\times n}$ is defined by $E_{lj} = 1$ if bus j is in balancing area $\mathcal{N}_l$ and $E_{lj} = 0$ otherwise. An edge $e \in \mathcal{E}$ is called a *tie-line* if its endpoints belong to different balancing areas in $\mathcal{P}^{\text{UC}}$ [5], [13]. As a result, the l-th row of $ECf = 0$ ensures that the branch flow deviations on the tie-lines connected to balancing area $\mathcal{N}_l$ sum to zero.

UC is designed so that its controller dynamics, combined with the system dynamics (1), form a variant of projected primal-dual algorithms to solve (2). It is shown in [3]–[7] that when the optimization problem (2) is feasible, under mild assumptions, the closed-loop equilibrium under UC is globally asymptotically stable and it is an optimal point of (2). Such an optimal point is unique (up to a constant shift of θ) if the cost functions $c_j(\cdot)$ are strictly convex. This means that, after a failure (cut set or not), the post-contingency system will be driven by UC to an optimal solution of (2) (under appropriate assumptions). We refer the readers to [3]–[7] for specific controller designs and analysis.

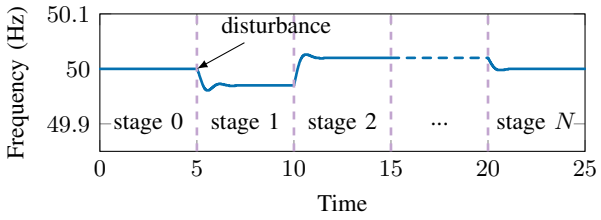


Fig. 1: An illustration of the failure propagation model.

We have introduced two distinct partitions of a power network so far: the tree partition $\mathcal{P}^{\text{tree}}$ in Part I and the balancing area partition $\mathcal{P}^{\text{UC}}$. In general, $\mathcal{P}^{\text{tree}}$ and $\mathcal{P}^{\text{UC}}$ can be different. However, when they do coincide, the underlying power grid inherits analytical properties from both tree partition and UC, making the system particularly robust against failures. Our proposed control strategy leverages precisely this connection, as we present in more detail in Section III.

To create a tree partition whose regions coincide with the balancing areas over which UC operates, we may have to switch off some tie-lines of $\mathcal{P}^{\text{UC}}$. The selection of these tie-lines can be systematically optimized, e.g., to minimize line congestion or inter-area flows on the resulting network; see [14] for more details. We henceforth assume that $\mathcal{P}^{\text{tree}} = \mathcal{P}^{\text{UC}}$, which means that the bridges and the tie-lines of the power network $\mathcal{G}$ coincide.

C. Failure Propagation

In full generality, the control strategy that we introduce applies to both generator failures and line failures. However, to simplify presentation, we focus on line failures in this paper. Recall from Part II that we describe the cascading failure process by keeping track of the set of outaged lines $\mathcal{B}(\mathfrak{n})$ over $\mathfrak{n} \in \{1, 2, \ldots, N\}$ at steady state. Instead of simply assuming that power flows redistribute according to the DC model with the same injections, we now assume that the system evolves at a fast timescale according to the frequency dynamics (1) during transient and converges to a closed-loop equilibrium. This equilibrium can be interpreted as an optimal solution of a certain optimization problem. For instance, if UC is adopted as the frequency control method, the equilibrium will be the optimal solution of (2). At the new closed-loop equilibrium, overloaded lines are tripped and the cycle repeats, as illustrated in Fig. 1. The crux of our failure propagation model lies in the interplay between the slow timescale line tripping process and the fast-timescale dynamics.

More specifically, for each stage $\mathfrak{n} \in \{1, 2, \ldots, N\}$, the system evolves according to the dynamics (1) on the topology $\mathcal{G}(\mathfrak{n})$, and converges to an equilibrium point $x^*(\mathfrak{n}) = (\theta^*(\mathfrak{n}), \omega^*(\mathfrak{n}), d^*(\mathfrak{n}), f^*(\mathfrak{n}))$ that solves an optimization problem over $\mathcal{G}(\mathfrak{n})$. If all the branch flows $f^*(\mathfrak{n})$

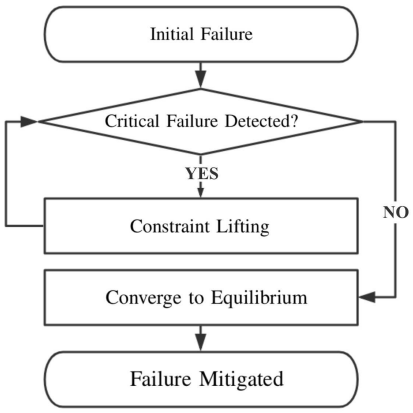


Fig. 2: Flowchart of the events after an initial failure under the proposed control strategy.

are within the corresponding line limits at equilibrium, then $x^*(\mathfrak{n})$ is a secure operating point and the cascade stops. Otherwise, let $\mathcal{F}(\mathfrak{n})$ be the subset of lines whose branch flows exceed the corresponding line limits. The lines in $\mathcal{F}(\mathfrak{n})$ operate above their limits in steady state, so we assume they trip at the end of stage $\mathfrak{n}$; i.e., $\mathcal{B}(\mathfrak{n}+1) = \mathcal{B}(\mathfrak{n}) \cup \mathcal{F}(\mathfrak{n})$. Line overloads during the transient phase before the system converges to $x^*(\mathfrak{n})$ are considered tolerable because the transient dynamics in (1) does not last long enough to overheat a line [5] (spanning only seconds to a few minutes). This process then repeats for the subsequent stages.

Definition 2. *Given a cascading failure process described by $\mathcal{B}(\mathfrak{n})$, with $\mathfrak{n} \in \{1, 2, \ldots, N\}$, the set $\mathcal{B}(1)$ is said to be its* **initial failure***. An initial failure $\mathcal{B}(1)$ is said to be* **critical** *if the UC optimization* (2) *is infeasible over $\mathcal{G}(1) := (\mathcal{N}, \mathcal{E}\backslash\mathcal{B}(1))$, or* **non-critical** *otherwise.*

To formally state our localization result, we define the following concept to clarify the precise meaning of a region being "local" with respect to an initial failure.

Definition 3. *Given an initial failure $\mathcal{B}(1)$, we say that a tree-partition region $\mathcal{N}_l$ is* **associated** *with $\mathcal{B}(1)$ if there exists an edge $e = (i, j) \in \mathcal{B}(1)$ such that either $i \in \mathcal{N}_l$ or $j \in \mathcal{N}_l$.*

As we discuss below, our control strategy possesses a strong localization property for both non-critical and critical failures in the sense that only the operation of the associated regions are adjusted whenever possible.

III. Proposed Control Strategy: Summary

Our strategy consists of two phases: a planning phase in which a tree partition of the network is created and an operation phase in which UC actively monitors and autonomously reacts to line failures as part of its normal operation. Fig. 2 illustrates the sequence of events after an initial failure.

A. Planning Phase: Tree Partitioning of Balancing Areas

Each balancing area of a multi-area power network is managed by an independent system operator (ISO). Although these areas exchange power with each other as prescribed by economic dispatch, their operations are relatively independent and it is desirable to ensure that system disturbances in one area do not have a significant impact on the other areas. This is usually achieved via the zero area control error constraint in secondary frequency control [13], and is enforced in UC with (2e). As mentioned in Section II, such balancing areas typically do not form a tree partition of the transmission network, as redundant lines are believed to be critical in maintaining $N-1$ security of grid [13], [15], [16].

We propose to create a tree partition whose regions coincide with the balancing areas over which UC operates. This can be done by switching off a subset of the inter-area tie-lines so that the resulting reduced graph (see Section IV in Part I [1]) forms a tree. The switching actions only need to be carried out in the planning phase, as line failures that occur during the operating phase do not affect the tree partition already in place.[2] It is interesting to notice that when the subset of lines to switch off is chosen carefully, tree partitioning not only localizes the impact of line failures, but can also *improves* overall reliability. This seemingly counter-intuitive phenomenon is illustrated by our case studies in Section VI-A.

B. Operating Phase: Extending Unified Controller

Once a tree partition is formed, the power network under UC operates as a closed-loop system and responds to disturbances such as line failures or loss of generator/load in an autonomous manner. In normal conditions where the system disturbances are small, UC always drives the power network back to an equilibrium point that can be interpreted as an optimal solution of (2). This is the case, for instance, when non-critical failures (see Definition 2) happen, and therefore such failures are always properly mitigated.

However, in extreme scenarios where a major disturbance (e.g., a critical failure) affects the system, the optimization problem (2) that UC aims to solve can be infeasible. In other words, it is physically impossible for UC to achieve all of its control objectives after such a disturbance. This makes UC unstable (see Proposition 6) and, may lead to successive failures. There is therefore a need to extend the version of UC proposed in [3]–[7] with two features: (a) a detection mechanism that monitors the system state and detects critical failures promptly; and (b) a constraint lifting mechanism that responds to critical failures by proactively relaxing certain constraints of (2) to ensure system stability can be reached at minimal cost.

Our technical results in Section V-A suggest a way to implement both components as part of the normal operation of UC. System operators can prioritize different balancing areas by specifying the sequence of constraints to lift in response to extreme events. This allows the non-associated regions to be progressively involved and coordinated in a systematic fashion when mitigating critical failures. We discuss some potential schemes in Section V-B.

C. Guaranteed Mitigation and Localization

We show in detail in Sections IV and V that the proposed strategy provides strong guarantees in the mitigation and localization for both non-critical and critical failures. Specifically, it ensures that the cascading process is always

[2] In fact, line failures can lead to "finer" tree partitions as more regions are potentially created when lines are removed from service.

stopped (a) after a non-critical failure by the associated regions, and the operating points of non-associated regions are not impacted in equilibrium; and (b) after a critical failure when constraints in (2) are lifted in a progressive manner specified by the system operator. Thus the proposed strategy can always prevent successive failures, while localizing the impact of the initial failures as much as possible.

IV. Localizing Non-critical Failures

In this section, we consider non-critical failures, as defined in Section II, and prove that such failures are always fully mitigated within the associated regions.

We first characterize how the system operating point shifts in response to such failures. Recall that if an initial failure $\mathcal{B}(1)$ is non-critical, the UC optimization (2) is feasible and thus the new operating point $x^*(1) := (\theta^*(1), \omega^*(1), d^*(1), f^*(1))$ satisfies all the constraints in (2). In particular, none of the line limits is violated at $x^*(1)$ by (2f), i.e. $x^*(1)$ is a secure operating point and the cascade stops ($\mathcal{F}(1) = \emptyset$). Moreover the power flows on bridges remain unchanged in equilibrium from their pre-contingency values, as the next result says. Note that all branch flows may deviate from their pre-contingency values during transient.

Lemma 4. *Given a non-critical initial failure $\mathcal{B}(1)$, the new operating point $x^*(1)$ prescribed by the UC satisfies $f_e^*(1) = 0$ for every bridge e.*

This lemma, proved in Appendix I, shows that tree partitioning enables UC to achieve more than what it is originally designed for in [3]–[7]: the extended UC not only enforces zero area control errors through (2e), it also guarantees zero flow deviations on all tie-lines.

The following proposition is another result of this type, which clarifies how tree partitioning induces a localization property in UC. See Appendix II for a proof.

Proposition 5. *Assume $c_j(\cdot)$ is strictly convex and achieves its minimum at $d_j = 0$ for all $j \in \mathcal{N}$. Given a non-critical initial failure $\mathcal{B}(1)$, if a tree-partition region $\mathcal{N}_l$ is not associated with $\mathcal{B}(1)$, then at equilibrium $x^*(1)$ we have $d_j^*(1) = 0$ for all $j \in \mathcal{N}_l$.*

The intuition of this proposition is easy to explain: Lemma 4 implies that the tie-line flows, which are the only coupling among regions, are zero in equilibrium; thus the UC optimization (2) over different regions are "decoupled" and hence the operating points of non-associated regions should remain unchanged. Such "decoupling", however, is *only* achieved over a tree partition; that is, if the balancing areas that UC operates over do not form a tree partition, Proposition 5 may not hold even when all tie-line flows remain unchanged.

As an example, consider the double-ring network introduced in Part I [1] as shown in Fig. 3. If we put the left and right rings as two balancing areas, and enforce the flow deviations on the two tie lines (G, L) and (G', L') to be zero, then the phase angle differences $\theta_G - \theta_L$ and $\theta_{G'} - \theta_{L'}$ also remain unchanged. As a result, we see from

$$\theta_G - \theta_{G'} = (\theta_G - \theta_L) + (\theta_L - \theta_{L'}) + (\theta_{L'} - \theta_{G'}) = \theta_L - \theta_{L'}$$

that $\theta_G - \theta_{G'}$ is fully determined by $\theta_L - \theta_{L'}$. In other words, the branch flow over (G, G') depends on the branch flow over (L, L') and, hence, the two regions cannot be decoupled.

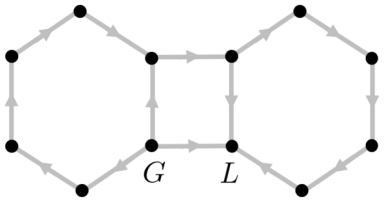


Fig. 3: A double-ring network.

This example suggests that a rigorous proof for Proposition 5 is more technical and involved. We refer the readers to Appendix II for how such a subtlety on region "separation" is precisely resolved by Lemma 15 from Part I [1] that relates the solution space of $L = CBC^T$ to tree partitions.

Two remarks are in order. First, Proposition 5 reveals that with the proposed control strategy, when the system converges to an equilibrium after a non-critical failure, the injections and power flows in non-associated regions remain unchanged, even though they fluctuate during transient according to (1). Our control scheme guarantees that non-critical failures in a balancing area do not impact the operations of other areas, achieving stronger balancing area independence than that ensured by zero area control error alone.

Second, unlike the scheme in Part II [2], cut set failures are treated in exactly the same way as non-cut failures provided that they are non-critical. Furthermore, the impact of a cut set failure is localized to the associated regions. This is in stark contrast with the global impact of a bridge failure in Part II [2] and is the key benefit of integrating UC with tree partitioning.

V. Controlling Critical Failures

We now consider the case where the initial failure is critical. This may happen when a major generator or transmission line is disconnected from the grid.

A. Unified Controller under Critical Failures

Since UC is a concept that has emerged from the frequency regulation literature, the underlying optimization (2) is always assumed to be feasible in existing studies [3]–[7]. As such, little is known about the behavior of UC if this assumption is violated when a critical failure happens. We now characterize the limiting behavior of UC in this setting.

In order to do this, we first formulate the exact controller dynamics of UC. Unfortunately, there is no standard way to do this as multiple designs of UC have been proposed in the literature [3]–[7], each with its own strengths and weaknesses. Nevertheless, all of the proposed controller designs are (approximately) projected primal-dual algorithms for the optimization problem (2) satisfying two assumptions that we now state. Let λ_i, for $i \in \left\{1, 2, \cdots, n + 3m + \left|\mathcal{P}^{\text{UC}}\right|\right\}$, be the dual variables corresponding to the constraints (2c)-(2f).

UC1: For all $j \in \mathcal{N}$, $\underline{d}_j \le d_j(t) \le \overline{d}_j$ is satisfied for all t. This is achieved either via a projection operator that maps $d_j(t)$ to this interval or by requiring the cost function $c_j(\cdot)$ to approach infinity near these boundaries.

UC2: The primal variables f, θ and the dual variables λ_i are updated by a primal-dual algorithm[3] to solve (2).

[3]We do not consider the specific variants of primal-dual algorithms that are proposed in different designs of UC, since the standard primal-dual algorithm is often a good approximation.

Proposition 6. *Assume UC1-UC2 hold. If* (2) *is infeasible, then there exists a dual variable* λ_i *such that:*

$$\limsup_{t\to\infty} |\lambda_i(t)| = \infty.$$

See Appendix III for a proof. Proposition 6 implies that, after a critical failure, UC cannot drive the system to a proper and safe operating point. It is a form of instability in the sense that at least one dual variable will take arbitrarily large values, and suggests a way to detect critical failures. Specifically, since Proposition 6 guarantees that at least one dual variable becomes arbitrarily large in UC operation when (2) is infeasible, we can set thresholds for the dual variables and raise an infeasibility warning if any of them exceeds their thresholds. By doing so, critical failures can be detected in a distributed fashion during normal operation of UC.

Since non-critical failures may also cause relatively large dual variable values in transient states, the choice of the thresholds inevitably involves tradeoffs. Tighter thresholds allow critical failures to be detected more promptly, yet also lead to a larger false alarm rate. In practice, these thresholds should be chosen carefully by the operator in accordance to specific system parameters and application scenarios.

B. Constraint Lifting as a Remedy

In the event of a critical failure, it is impossible for UC to simultaneously achieve all of its control objectives and constraints. This can lead to instability and thus successive failures. We can mitigate this by progressively lifting certain constraints from UC in two different ways without compromising the basic objective of stabilizing the system:

- The zero area control error constraints (2e) between specific pairs of balancing areas can be lifted. In practice, this means the controller now involves more balancing areas in failure mitigation.
- Loads can be shedded, which is reflected in (2) by enlarging the range $[\underline{d}_j, \overline{d}_j]$ for corresponding load buses.

By iteratively lifting the two types of constraints above, one can guarantee the feasibility of (2) and ensure that the system converges to a stable point that is free from successive failures. This, however, comes with the cost of potential load loss, and thus must be carried out judiciously.

The iterative relaxation procedure can follow predetermined rules specified by the system operator to prioritize different objectives. As an example, one can minimize load loss by relaxing possibly all area control error constraints before relaxing injection bounds on load buses. This will utilize all the contingency and regulation reserves globally across all regions to meet demand before shedding load as a last resort. In contrast, if the localization of failure impact should be prioritized, the operator can choose to first lift load injection bounds in the associated areas and then progressively lift area control error constraints to get more balancing areas involved.

VI. Case Studies

In this section, we evaluate the performance of the proposed control strategy on the IEEE 118-bus and IEEE 2736-bus (the Polish network) test systems, with respect to $N-k$ security standard and localization performance under different levels of system congestion.

A. $N-k$ Security under Different Congestion Levels

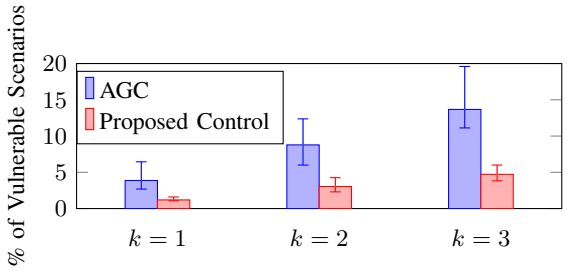
We first focus on system robustness with respect to $N-k$ security standard on the IEEE 118-bus system. This test network has two balancing areas shown as Region 1 and Region 2 in Appendix IV. To form a tree partition, three lines $(15, 33)$, $(19, 34)$, and $(23, 24)$ are switched off and this new network is referred as the revised network in sequel.

We compare UC on the tree-partitioned revised network and classical AGC on the *original* network. UC is modeled by the optimization problem (2) and AGC is modeled by (2) without the line limits (2f). A failure scenario is said to be *vulnerable* if the initial failure leads to successive failures or loss of load. To compare the performance between our proposed approach and AGC, we collect statistics on (a) vulnerable scenarios as a percentage of the total simulated scenarios, and (b) load loss rate (LLR) which is defined as the ratio of the total load loss to the original total demand. We do not perform time-domain simulations, but assume the closed-loop systems under UC and AGC converge to their respective equilibrium points that solve corresponding optimizations respectively.

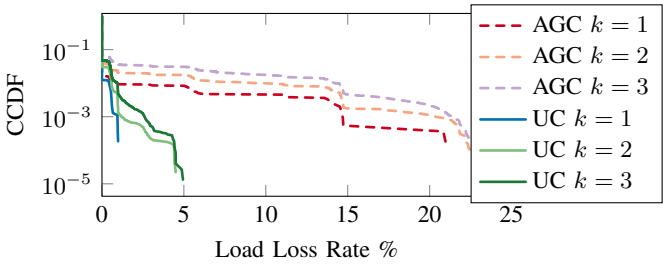
The failure scenarios are created as follows. First, we generate a variety of load injections (as summarized in Table IV in Appendix V) by adding random perturbations to the nominal load profile from [17] and then solve the DC OPF to obtain the corresponding generator operating points. Second, we sample over the collection of all subsets that consist of k transmission lines of the IEEE 118-bus test network. Finally, for each sampled subset of k lines, we remove all lines in this subset as initial failure and simulate the cascading process thus triggered. Our simulations cover the cases $k = 1, 2, 3$, resulting in roughly 138,600 failure scenarios.

Fig. 4(a) shows the average, minimum, and maximum percentage of vulnerable scenarios across all sampled failure scenarios, while Fig. 4(b) plots the complementary cumulative distribution (CCDF) of the load loss rates. The simulation results show that the proposed control incurs both substantially fewer vulnerable scenarios and much less loss of load in all cases compared to AGC. This difference is particularly pronounced when multiple lines are tripped simultaneously ($k = 2, 3$). We highlight that in our simulations, UC operates over the tree-partitioned network (while AGC operates over the original network) in which some of the tie-lines are switched off and hence some transfer capacity is removed from the system. Moreover, the newly created bridge $(30, 38)$ in the tree partition is never vulnerable under the proposed control in all the scenarios we have studied.

We then illustrate the improvement of the proposed approach over AGC under different congestion levels. To do so, we scale down the line capacities to $\alpha = 0.9, 0.8, 0.7$ of the base values and collect statistics for all single line initial failures ($k = 1$). Our results are summarized in Fig. 5, which again show that the proposed approach significantly outperforms AGC in all scenarios, especially those in which the system is congested. Again, the bridge $(30, 38)$ in the tree partition is never vulnerable under the proposed control in these scenarios.

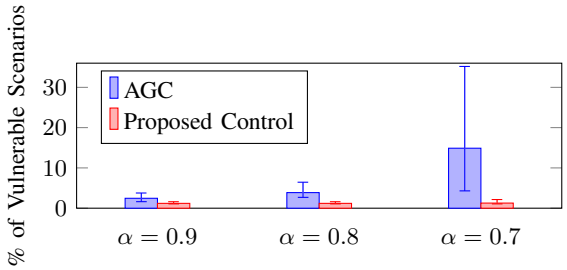


(a) Average fraction of vulnerable scenarios for $k = 1, 2, 3$.

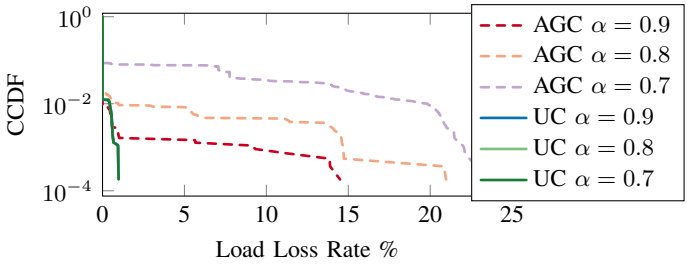


(b) CCDF for load loss rate for $k = 1, 2, 3$.

Fig. 4: System robustness in terms of the $N - k$ security standard.



(a) Average fraction of vulnerable scenarios for different congestion levels.



(b) CCDF for load loss rate for different congestion levels.

Fig. 5: System robustness under different levels of congestion obtained scaling line capacities by the factor α.

B. Localized Failure Mitigation

In this subsection, we consider a specific constraint lifting rule that progressively involves other regions by relaxing area control error constraints only if local load shedding within the associated regions is not enough to make problem (2) feasible. This rule prioritizes localization of the initial failure. With this rule implemented, we show that the proposed control strategy can localize cascading failures within the associated regions with negligible load loss. The experiments are carried out over two networks: (a) a further-decomposed tree partition of the IEEE 118-bus test network, and (b) a much larger-scale Polish network consisting of 2736 buses and 3504 transmission lines.

For the IEEE 118-bus test network, we switch off 4 additional lines, which refines the tree partition used in the previous subsection since it further decomposes Region 2 into two balancing areas (as shown in Fig. 7 in Appendix IV). The generator capacities are scaled down by 60% so that the total generation reserve is roughly 20%. We create different congestion levels by scaling the line capacities according to a factor $\alpha = 0.9, 0.8, 0.7$ and iterate over all single transmission lines as initial failures. The injections are the same as that for the $N - 1$ test in the previous subsection.

The statistics on the fraction of vulnerable scenarios and LLRs for this experiment are summarized in Table II. We observe that the proposed control strategy never incur more than 2.21% LLR across all tested injections and congestion levels. Furthermore, for this specific network, the proposed approach localizes *all* failures to the associated regions, i.e., the tie-line constraints are never lifted. This localization phenomenon can more clearly be noticed in Fig. 6(a), which shows the CCDF of the number of generators whose operating points are adjusted in response to the initial failures. The majority of failures lead to operating point adjustments

TABLE II: Statistics on failure localization over the IEEE 118-bus test network.

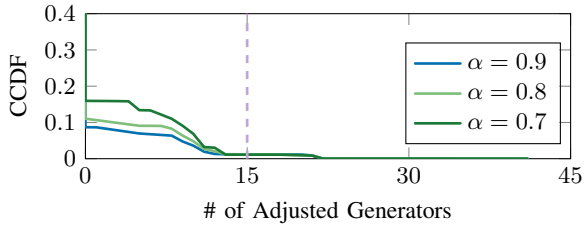
Line Capacity	$\alpha = 0.9$	$\alpha = 0.8$	$\alpha = 0.7$
Avg. % of Vul. Sce.	3.53	3.68	3.82
Avg. (Max.) LLR(%)	0.55 (1.06)	0.56 (2.17)	0.59 (2.21)

on less than 15 generators, which is roughly the number of generators within a single region. The small portion of failures that impact more than 15 generators are bridge failures, which by definition have two associated regions and thus more "local" generators.

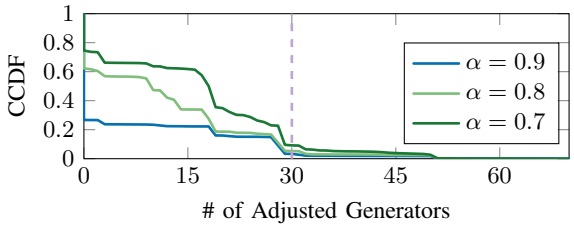
For the Polish network, we switch off 78 transmission lines from the original network, creating a tree partition with 4 regions of 1430, 818, 359 and 129 buses respectively. Similar to the setup for the IEEE-118 test network, the generation capacities are scaled properly so that the total generation reserve is roughly 20%, and the line capacities are scaled down to $\alpha = 0.9, 0.8, 0.7$ to create different congestion levels. We then iterate over all single line failures and the statistics from our experiments are summarized in Table III. Our results show that for this test network, more than 86% of the single line failures can be mitigated locally within a single region for all congestion levels. In addition, the worst case LLR is roughly 3% across all simulated scenarios, with an average that is no higher than 0.07%. Similar to the IEEE 118-bus test network, the number of generators whose operating points are adjusted by the proposed control strategy in response to the failures is small, as shown in Fig. 6(b), confirming failure localization.

VII. Conclusion

Across Part I, II, and III of this paper, we have proposed a new approach to grid reliability. It provides strong analytical



(a) CCDF for generator response on IEEE 118-bus network.



(b) CCDF for generator response on IEEE Polish network.

Fig. 6: Generator response over IEEE 118-bus and Polish network.

TABLE III: Statistics on failure localization over the Polish network.

Line Capacity	$\alpha = 0.9$	$\alpha = 0.8$	$\alpha = 0.7$
Scenarios Mitigated with one Region (%)	92.39	88.63	86.91
Scenarios Mitigated with 2-3 Regions (%)	6.44	9.48	10.40
Scenarios Mitigated with All Regions (%)	1.17	1.89	2.69
Avg. (Max.) LLR(%)	0.05 (2.93)	0.05 (2.94)	0.07 (3.24)
Avg. # of Gen. Adj.	6.52	11.66	16.37

guarantees of both the localization and mitigation of cascading failures. In Part I, we establish a fundamental theory and analytical characterizations of power systems from a spectral perspective and introduce the concept of tree partition. We then demonstrate in Part II that the tree partitioning network structure provides a precise characterization of line failure localizability. Finally in Part III, we integrate tree partitioning and unified controller for frequency regulation to mitigate line failures in real time. Our case studies on the IEEE 118-bus and 2736-bus test systems show that the proposed control scheme can greatly improve overall reliability compared to the current practice. In particular, the new control prevents successive failures from happening while localizing the impacts of initial failures. When load shedding is inevitable, the proposed strategy incurs significantly less load loss.

This work can be extended in several directions. First, our model builds upon linearized swing and power flow dynamics, which are accurate for small deviations but less so under large disruptions. It is thus crucial to understand how the non-linearity and large deviations impact our results. Second, the proposed control strategy may require certain tie-lines to be switched off to create a finer tree partition. It would be useful to optimize the selection of these lines (see [14] for more details). Third, in addition to power flow redistribution, line capacities also play important roles in the cascading process, but our current theory does not incorporate them in a way that is convenient for analysis. It is important to investigate how adjustments on line capacities can be incorporated in our framework to further improve system reliability. Finally, it will be important to understand the tradeoff between grid redundancy and tree partitioning in achieving optimal grid reliability. Adding redundancy helps maintain connectivity and transfer capacity in line failure events, while tree partitioning can localize failure propagation and reduce the risk of large-scale blackouts. A good solution likely needs to take both aspects into account and locate a sweet spot in between.

REFERENCES

[1] L. Guo, C. Liang, A. Zocca, S. H. Low, and A. Wierman, "Localization & Mitigation of Cascading Failures in Power Systems, Part I: Spectral Representation & Tree partition," *arXiv:2005.10199*.

[2] ——, "Localization & Mitigation of Cascading Failures in Power Systems, Part II: Localization," *arXiv:2005.11320*.

[3] C. Zhao, U. Topcu, N. Li, and S. Low, "Design and stability of load-side primary frequency control in power systems," *Automatic Control, IEEE Transactions on*, vol. 59, no. 5, pp. 1177–1189, 2014.

[4] N. Li, C. Zhao, and L. Chen, "Connecting automatic generation control and economic dispatch from an optimization view," *IEEE Transactions on Control of Network Systems*, vol. 3, no. 3, pp. 254–264, 2016.

[5] C. Zhao, E. Mallada, S. Low, and J. Bialek, "A unified framework for frequency control and congestion management," in *Power Systems Computation Conference (PSCC), 2016*. IEEE, 2016, pp. 1–7.

[6] E. Mallada, C. Zhao, and S. H. Low, "Optimal load-side control for frequency regulation in smart grids," *IEEE Transactions on Automatic Control*, vol. 62, no. 12, pp. 6294–6309, 12 2017. [Online]. Available: https://mallada.ece.jhu.edu/pubs/2017-TAC-MZL.pdf

[7] C. Zhao, E. Mallada, S. H. Low, and J. Bialek, "Distributed plug-and-play optimal generator and load control for power system frequency regulation," *International Journal of Electrical Power & Energy Systems*, vol. 101, pp. 1–12, 2018.

[8] J. Jung, C.-C. Liu, S. L. Tanimoto, and V. Vittal, "Adaptation in load shedding under vulnerable operating conditions," *IEEE Trans. Power Systems*, vol. 17, no. 4, pp. 1199–1205, November 2002.

[9] R. Faranda, A. Pievatolo, and E. Tironi, "Load shedding: A new proposal," *IEEE Trans. Power Systems*, vol. 22, no. 4, pp. 2086–2093, November 2007.

[10] D. Bienstock, "Optimal control of cascading power grid failures," in *CDC*, Dec 2011, pp. 2166–2173.

[11] S. Pahwa, C. Scoglio, S. Das, and N. Schulz, "Load shedding strategies for preventing cascading failures in power grid," *Electric Power Components and Systems*, July 2013.

[12] Q. Ba and K. Savla, "Computing optimal control of cascading failure in DC networks," *IEEE Trans. Automatic Control*, 2019, accepted for publication.

[13] A. R. Bergen, *Power systems analysis*. Pearson Education India, 2009.

[14] A. Zocca, L. Guo, C. Liang, A. Wierman, and S. H. Low, "A spectral representation of power systems with applications to adaptive partitioning, failure localization, and network optimization," *Manuscript in preparation*.

[15] D. Bienstock and S. Mattia, "Using mixed-integer programming to solve power grid blackout problems," *Discrete Optimization*, vol. 4, no. 1, pp. 115 – 141, 2007.

[16] P. Hines, S. Talukdar *et al.*, "Controlling cascading failures with cooperative autonomous agents," *International journal of critical infrastructures*, vol. 3, no. 1, p. 192, 2007.

[17] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE TPS*, vol. 26, no. 1, pp. 12–19, 2011.

[18] S. Soltan, D. Mazauric, and G. Zussman, "Analysis of failures in power grids," *IEEE TCNS*, no. 99, 2015.

[19] J. Yan, Y. Tang, H. He, and Y. Sun, "Cascading failure analysis with DC power flow model and transient stability analysis," *IEEE TPS*, vol. 30, no. 1, pp. 285–297, 2015.

[20] A. Bernstein, D. Bienstock, D. Hay, M. Uzunoglu, and G. Zussman, "Power grid vulnerability to geographically correlated failures: Analysis and control implications," in *IEEE INFOCOM*, 2014, pp. 2634–2642.

APPENDIX I
PROOF OF LEMMA 4

To simplify the notation, we drop the stage index (1) from x^* and denote $x^* = (\theta^*, \omega^*, d^*, f^*)$. Given a bridge $e = (j_1, j_2)$ of $\mathcal{G}$, removing e from $\mathcal{G}$ partitions $\mathcal{G}$ into two connected components, say $\mathcal{C}_1$ and $\mathcal{C}_2$. Without loss of generality, assume $j_1 \in \mathcal{C}_1$ and $j_2 \in \mathcal{C}_2$. For a region $\mathcal{N}_v$ from $\mathcal{P}$, we say $\mathcal{N}_v$ is within $\mathcal{C}_1$ if for any $j \in \mathcal{N}_v$ we have $j \in \mathcal{C}_1$. It is easy to check from the definition of tree partitions that any region $\mathcal{N}_v$ from $\mathcal{P}$ is either within $\mathcal{C}_1$ or within $\mathcal{C}_2$, and e is the only edge in $\mathcal{G}$ that has one endpoint in $\mathcal{C}_1$ and the other endpoint in $\mathcal{C}_2$.

Let $\mathcal{P}'$ be the set of regions within $\mathcal{C}_1$ from $\mathcal{P}$, and let $\mathbf{1}_{\mathcal{P}'} \in \{0,1\}^{|\mathcal{P}|}$ be its characteristic vector (that is, the l-th component of $\mathbf{1}_{\mathcal{P}'}$ is 1 if $\mathcal{N}_l \in \mathcal{P}'$ and 0 otherwise). Given two buses i and j, we denote $i \to j$ if $(i,j) \in \mathcal{E}$ and $j \to i$ if $(j,i) \in \mathcal{E}$. With these notations, from (2e), we have

$$\begin{aligned} 0 &= \mathbf{1}_{\mathcal{P}'}^T ECf^* \\ &= \sum_{l:\mathcal{N}_l \in \mathcal{P}'} \sum_{i \in \mathcal{N}_l} \left(\sum_{j:j\to i} f_{ji}^* - \sum_{j:i\to j} f_{ij}^* \right) \\ &= \sum_{i:i\in\mathcal{C}_1} \left(\sum_{j:j\to i} f_{ji}^* - \sum_{j:i\to j} f_{ij}^* \right) \\ &= f_e^* + \sum_{i:i\in\mathcal{C}_1} \left(\sum_{j:j\to i, j\in\mathcal{C}_1} f_{ji}^* - \sum_{j:i\to j, j\in\mathcal{C}_1} f_{ij}^* \right) \end{aligned} \quad (3)$$

where (3) follows because the only edge with one endpoint in $\mathcal{C}_1$ and the other endpoint in $\mathcal{C}_2$ is e. Note that

$$\begin{aligned} 0 &= \sum_{(i,j)\in\mathcal{E}_1} \left(f_{ij}^* - f_{ij}^* \right) \\ &= \sum_{i:i\in\mathcal{C}_1} \left(\sum_{j:j\to i, j\in\mathcal{C}_1} f_{ji}^* - \sum_{j:i\to j, j\in\mathcal{C}_1} f_{ij}^* \right) \end{aligned}$$

where $\mathcal{E}_1$ is the set of edges with both endpoints in $\mathcal{C}_1$. From (3), we see that $f_e^* = 0$. Since the bridge e is arbitrary, we have thus proved the desired result. □

APPENDIX II
PROOF OF PROPOSITION 5

To simplify notation, we drop the stage index (1) from the equilibrium x^* and denote $x^* = (\theta^*, \omega^*, d^*, f^*)$, $p^* = r + d^*$.

First, we construct a different point $\tilde{x}^* = (\tilde{\theta}^*, \tilde{\omega}^*, \tilde{d}^*, \tilde{f}^*)$ by changing certain entries of x^* as follows: (a) replace d_j^* with $\tilde{d}_j^* = 0$ for all $j \in \mathcal{N}_l$; (b) replace f_e^* with $\tilde{f}_e^* = 0$ for $e \in \mathcal{E}$ that have both endpoints in $\mathcal{N}_l$; and (c) replace θ^* by a solution $\tilde{\theta}^* = L^\dagger \tilde{p}^*$ obtained from solving the DC power flow equations with injections $\tilde{p}^* = r + \tilde{d}^*$. All other entries of x^* remain unchanged in $\tilde{x}^*$. Since $c_j(\cdot)$ attains its minimum at $d_j = 0$, $\tilde{x}^*$ achieves at most the same objective value (2a) as x^*. Thus $\tilde{x}^*$ must be an optimal point of (2), provided it is feasible.

When the cost functions $c_j(\cdot)$ are strictly convex, the optimal solution to (2) is unique in d^* and f^* (θ^* is also unique up to an arbitrary reference angle). As a result, if the constructed point $\tilde{x}^*$ is feasible, We can then conclude that $\tilde{x}^* = x^*$ (up to an arbitrary reference angle).

We now prove the feasibility of $\tilde{x}^*$. The construction of $\tilde{x}^*$ ensures that (2e)(2f)(2g) are satisfied. If we can show that $\tilde{f}^* = BC^T\tilde{\theta}^*$, then since $\tilde{\theta}^*$ is obtained by solving the DC power flow equations from $CBC^T\tilde{\theta}^* = \tilde{p}^*$, constraints (2c) and (2d) are also satisfied, proving the feasibility of $\tilde{x}^*$. It thus suffices show $\tilde{f}^* = BC^T\tilde{\theta}^*$. To do so, we first establish the following lemma:

Lemma 7. *For any tree-partition region $\mathcal{N}_z$ in $\mathcal{P}$, we have*

$$\sum_{j\in\mathcal{N}_z} p_j^* = \sum_{j\in\mathcal{N}_z} \tilde{p}_j^* = 0.$$

Proof. Let $\mathbf{1}_{\mathcal{N}_z} \in \mathbb{R}^{|\mathcal{N}|}$ be the characteristic vector of $\mathcal{N}_z$, that is, the j-th component of $\mathbf{1}_{\mathcal{N}_z}$ is 1 if $j \in \mathcal{N}_z$ and 0 otherwise. Summing (2c) over $j \in \mathcal{N}_z$, we have:

$$\sum_{j\in\mathcal{N}_z} p_j^* = \mathbf{1}_{\mathcal{N}_z}^T Cf = (ECf)_z = 0,$$

where $(ECf)_z$ is the z-th row of ECf.

For $\mathcal{N}_l$, we have $\tilde{p}_j^* = 0$ for $j \in \mathcal{N}_l$ by construction and hence $\sum_{j\in\mathcal{N}_l} \tilde{p}_j^* = 0$. For $\mathcal{N}_z$ that is different from $\mathcal{N}_l$, we have $\tilde{p}_j^* = p_j^*$ for any $j \in \mathcal{N}_z$ by construction. Thus, $\sum_{j\in\mathcal{N}_z} \tilde{p}_j^* = 0$, completing the proof. □

Consider now a region $\mathcal{N}_w$ that is different from $\mathcal{N}_l$. In this case, we do not change the injections from x^* when constructing $\tilde{x}^*$, thus $p_j^* - \tilde{p}_j^* = 0$ for all $j \in \mathcal{N}_w$. From Lemma 7, we see that $\sum_{j\in\mathcal{N}_z} \left(p_j^* - \tilde{p}_j^*\right) = 0$ for all z. Since both (p^*, θ^*) and $(\tilde{p}^*, \tilde{\theta}^*)$ satisfy the DC power flow equations, we have

$$CBC^T \left(\theta^* - \tilde{\theta}^*\right) \;=\; p^* - \tilde{p}^*.$$

By Lemma 15 of Part I, we then have $\theta_j^* - \tilde{\theta}_j^*$ is a constant over $\overline{\mathcal{N}}_w$, and thus

$$\tilde{\theta}_i^* - \tilde{\theta}_j^* \;=\; \theta_i^* - \theta_j^*$$

for all $i, j \in \overline{\mathcal{N}}_w$. This in particular implies

$$\tilde{f}_e^* \;=\; f_e^* \;=\; B_e(\theta_i^* - \theta_j^*) \;=\; B_e(\tilde{\theta}_i^* - \tilde{\theta}_j^*)$$

for all $e = (i,j)$ such that $i \in \mathcal{N}_w$ or $j \in \mathcal{N}_w$.

Finally, consider the region $\mathcal{N}_l$. We have $\tilde{p}_j^* = 0$ by construction. From Lemma 7 we have $\sum_{j\in\mathcal{N}_z} \tilde{p}_j^* = 0$ for all z. Thus by Lemma 15 of Part I and $CBC^T\tilde{\theta}^* = \tilde{p}^*$, we know $\tilde{\theta}_i^* = \tilde{\theta}_j^*$ for all $i, j \in \overline{\mathcal{N}}_l$. This implies that for any edge $e = (i,j)$ within $\mathcal{N}_l$, we have

$$\tilde{f}_e^* = 0 = B_e(\tilde{\theta}_i^* - \tilde{\theta}_j^*).$$

As a result, $\tilde{f}_e^* = B_e(\tilde{\theta}_i^* - \tilde{\theta}_j^*)$ for all $e \in \mathcal{E}$ and the proof is concluded. □

APPENDIX III
PROOF OF PROPOSITION 6

First, collect in the vector $x = (\theta, \omega, d, f) \in \mathbb{R}^{3n+m}$ all the decision variables of the UC optimization (2) and rewrite it in a more standard form as

$$\begin{aligned} \min_{\underline{d}\le d\le\overline{d}} \quad & c(d) && (4a) \\ \text{s.t.} \quad & Ax \le g && (4b) \\ & Cx = h, && (4c) \end{aligned}$$

where A, C, g, h are matrices (vectors) of proper dimensions from the optimization (2). Let λ_1, λ_2 be the corresponding dual variables to (4b) and (4c), respectively, and set $\lambda := [\lambda_1; \lambda_2]$ ($[\cdot;\cdot]$ here means matrix concatenation as a column). We can then write the Lagrangian for (4) as

$$\mathcal{L}(x, \lambda) = c(p) + \lambda_1^T (Ax - g) + \lambda_2^T (Cx - h).$$

By the assumption UC2, we know that:

$$\dot{\lambda}_1 = [Ax - g]^+_{\lambda_1} \tag{5a}$$
$$\dot{\lambda}_2 = Cx - h, \tag{5b}$$

where the projection operator $[\cdot]^+_{\lambda_1}$ is defined component-wise by

$$\left([x]^+_{\lambda_1}\right)_i := \begin{cases} x_i & \text{if } x_i > 0 \text{ or } (\lambda_1)_i > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Consider two closed convex sets $S_1 = \{x | Ax \le g, Cx = h\}$ and $S_2 = \{x | \underline{d} \le d \le \overline{d}\}$. If the optimization (2) is infeasible, then $S_1 \cap S_2 = \emptyset$. As a result, we can find a hyperplane that separates S_1 and S_2: there exists $q \in \mathbb{R}^{3n+m}, q_0 \in \mathbb{R}$ such that

$$q^T x > q_0, \forall x \in S_1 \text{ and } q^T x \le q_0, \forall x \in S_2.$$

This then implies the system

$$\begin{cases} Ax \le g \\ Cx = h \\ q^T x \le q_0 \end{cases}$$

is not solvable. By Farkas' Lemma, we can thus find vectors w_1, w_2, w_3 such that $w_1 \ge 0, w_3 \ge 0$ (the inequality is component-wise), $A^T w_1 + C^T w_2 + q w_3 = 0$, and $g^T w_1 + h^T w_2 + q_0 w_3 = -\epsilon < 0$.

Define $z = [w_1; w_2]$. We then see that under the UC controller, we have for any t:

$$\begin{aligned} z^T \dot{\lambda}(t) &= w_1^T [Ax(t) - g]^+_\lambda + w_2^T (Cx(t) - h) \\ &\ge w_1^T [Ax(t) - g]^+_\lambda + w_2^T (Cx(t) - h) \\ &\quad + w_3 (q^T x(t) - q_0) \qquad (7a) \\ &\ge w_1^T (Ax(t) - g) + w_2^T (Cx(t) - h) \\ &\quad + w_3 (q^T x(t) - q_0) \qquad (7b) \\ &= \left(A^T w_1 + C^T w_2 + q w_3\right) x(t) \\ &\quad - \left(w_1^T g + w_2^T h + w_3 q_0\right) \\ &= 0 + \epsilon > 0, \end{aligned}$$

where (7a) follows from $w_3 \ge 0$ and assumption UC1, which ensures $x(t) \in S_2$ and thus $q^T x(t) - q_0 \le 0$, and (7b) comes from $w_1 \ge 0$ and the fact that $[x]^+_\lambda \ge x$ for all x (the inequality is component-wise). Consequently,

$$z^T \lambda(t) - z^T \lambda(0) > \epsilon t$$

and thus

$$\lim_{t \to \infty} z^T \lambda(t) = \infty.$$

Finally, by noting

$$\lim_{t \to \infty} z^T \lambda(t) \le w_1^T \limsup_{t \to \infty} |\lambda_1(t)| + |w_2|^T \limsup_{t \to \infty} |\lambda_2(t)|,$$

the desired result follows. □

APPENDIX IV
ONE LINE DIAGRAM OF IEEE 118-BUS TEST NETWORK

The original IEEE 118-bus test network is shown by the solid lines and nodes in Fig. 7, in which the balancing areas are connected by multiple tie-lines, yielding a trivial tree-partition region as the whole network. For our experiments in Section VI-A, we switch off three lines $(15, 33)$, $(19, 34)$, and $(23, 24)$ to form a two-region tree partition, shown as Region 1 and Region 2 (which consists of Region 2(a) and Region 2(b)) in the diagram. For our experiments in Section VI-B, we further switch off the lines $(77, 82)$, $(96, 97)$, $(98, 100)$, $(99, 100)$ to create a three-region tree partition.

APPENDIX V
FAILURE SCENARIOS FOR $N - k$ SECURITY SIMULATION

To provide a comprehensive comparison between the proposed control strategy and AGC, we generate 136,000 failure scenarios (as summarized in Table IV) and simulate the corresponding cascading processes.

We adopt data from MatPower [17] as the nominal load profile, and add up to 25% random perturbations to the base value. For $k = 1, 2, 3$ initial line failures, we generate 100, 15, and 15 load profiles and further compute the optimal generation dispatch by DC OPF. For each load profile, we iterate over every single transmission line failure, and sample 3,000 and 5,000 failure scenarios for $k = 2, 3$ line failures respectively. In total, 136,000 failure scenarios are simulated for each control method.

TABLE IV: Simulation setup for $N - k$ security evaluation.

Case	$k = 1$	$k = 2$	$k = 3$
# of Load Profiles	100	15	15
# of Sampled Failures	186	3000	5000
Total Scenarios	18600	45000	75000

APPENDIX VI
RECOVERING PREVIOUS MODELS

The dynamic model (1) in Section II models secondary frequency control where the frequency deviations $\omega(t)$ are driven to zero. When we focus on controllers that only achieve primary frequency control, the equilibrium frequency ω^* may be nonzero. That is, as the system converges in this sense, the phase angles $\theta^*(t)$ do not necessarily stay at a constant value, but may change in constant rate over time. In such context, we can modify (1) as follows to describe primary frequency dynamics:

$$M_j \dot{\omega}_j = r_j + d_j - D_j \omega_j - \sum_{e \in \mathcal{E}} C_{je} f_e, \quad j \in \mathcal{N} \tag{8a}$$
$$f_{ij} = B_{ij}(\theta_i - \theta_j), \qquad (i, j) \in \mathcal{E}. \tag{8b}$$

By relaxing the requirement on $\omega^* = 0$ at equilibrium, the above model enables extra freedom in the choice of d_j. We now show that by using the classical droop control [13] as the dynamics for d_j's in (8), the cascading failure models from Part II and previous literature such as [18], [19] can be readily recovered. Indeed, as shown in [5], the *closed-loop* equilibrium of (8) under droop control is the unique[4]

[4]The equilibrium is unique up to an arbitrary reference phase angle.

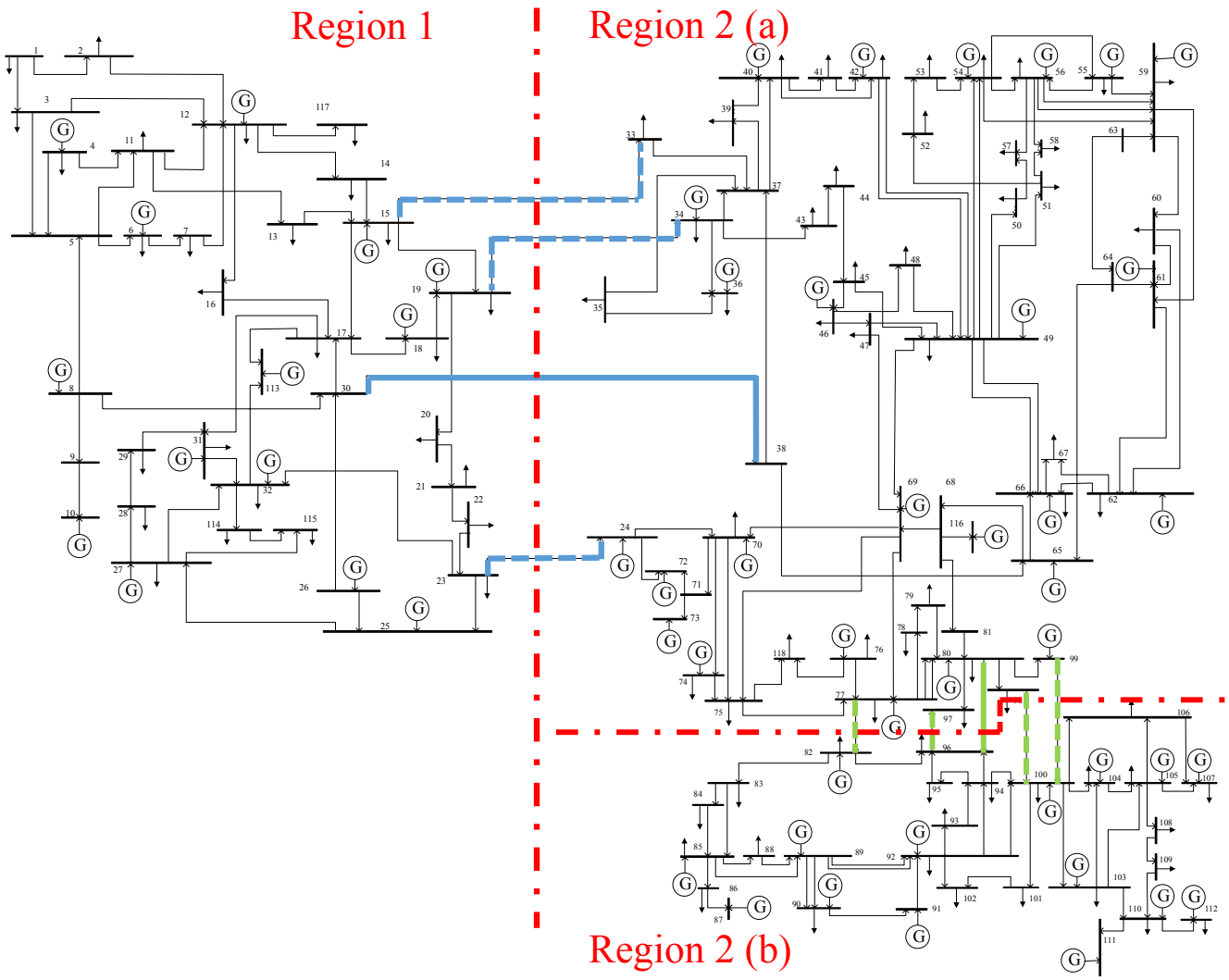


Fig. 7: One line diagram of the IEEE 118-bus test network with three balancing areas. Dashed blue lines are switched off to create a tree partition with 2 regions. Then dashed green lines are switched off to further partition Region 2 into 2 smaller regions. Solid blue and green lines are resulted bridges.

optimal solution to the following optimization on the post-contingency network:

$$\min_{\theta,\omega,d,f} \quad \sum_{j\in\mathcal{N}} \frac{d_j^2}{2Z_j} + \frac{D_j w_j^2}{2} \tag{9a}$$

$$\text{s.t.} \quad r - d - D\omega = Cf \tag{9b}$$

$$f - BC^T\theta = 0 \tag{9c}$$

$$\underline{p}_j \le r_j - d_j \le \overline{p}_j, \quad j \in \mathcal{N}, \tag{9d}$$

where Z_j's are the generators' participation factors [13]. By plugging (9c) into (9b), it is easy to check that any feasible point $x = (\theta, \omega, d, f)$ of (9) satisfies $\sum_j r_j = \sum_j (d_j + D_j\omega_j)$. Cauchy-Schwarz inequality then implies that

$$\left(\sum_{j\in\mathcal{N}} r_j\right)^2 = \left[\sum_{j\in\mathcal{N}} (d_j + D_j\omega_j)\right]^2 \le \sum_{j\in\mathcal{N}} \left(\frac{d_j^2}{2Z_j} + \frac{D_j\omega_j^2}{2}\right) \sum_{j\in\mathcal{N}} (2Z_j + 2D_j),$$

for which equality holds if and only if

$$d_j = \frac{Z_j}{\sum_j (Z_j + D_j)} \sum_j r_j, \quad \omega_j = \frac{\sum_j r_j}{\sum_j (Z_j + D_j)}. \tag{10}$$

Therefore, if the control limits (9d) are not active, (10) is always satisfied at the optimal point $x^* = (\theta^*, \omega^*, d^*, f^*)$.

Now, consider a line e being tripped from the transmission network $\mathcal{G}$, and for simplicity assume the control limits (9d) are not active. If e is a bridge, the tripping of e results in two islands of $\mathcal{G}$, say $\mathcal{D}_1$ and $\mathcal{D}_2$, and two optimization problems (9) correspondingly. For $l = 1, 2$, $\sum_{j\in\mathcal{D}_l} r_j$ represents the total net power imbalance in $\mathcal{D}_l$, and therefore (10) implies that droop control adjusts the system injections so that the power imbalance is distributed to all generators proportional to their participation factors in both $\mathcal{D}_1$ and $\mathcal{D}_2$. If $e = (i, j)$ is not a bridge, denoting the original flow on e before it is tripped as f_e, then $r_i = f_e$, $r_j = -f_e$ and $r_k = 0$ otherwise. As a result, we have $\sum_{j\in\mathcal{N}} r_j = 0$ in this case and thus (10) implies the system operating point remains unchanged in equilibrium, i.e., $d_j = \omega_j = 0, \forall j \in \mathcal{N}$. Moreover, one can show that this still holds when (9d) is active with a more

involved analysis on the KKT conditions of (9). This droop control mechanism recovers the failure propagation model in Part II and underlies some of previous results in the literature on cascading failures in power systems [18]–[20].