Adaptive Network Response to Line Failures in Power Systems

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

Abstract— **Transmission line failures in power systems propagate and cascade non-locally. In this work, we propose an adaptive control strategy that offers strong guarantees in both the mitigation and localization of line failures. Specifically, we leverage the properties of network *bridge-block decomposition* and a frequency regulation method called the *unified control*. If the balancing areas over which the unified control operates coincide with the bridge-blocks of the network, the proposed strategy drives the post-contingency system to a steady state where the impact of initial line outages is localized within the areas where they occurred whenever possible, stopping the cascading process. When the initial line outages cannot be localized, the proposed control strategy provides a configurable design that progressively involves and coordinates more balancing areas. 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 the majority of the simulated contingencies. Moreover, the proposed framework incurs significantly less load loss, if any, compared to AGC, in all our case studies.**

I. INTRODUCTION

Transmission line failures in power systems propagate both locally and non-locally, making it challenging to design control methods that reliably prevent and control line failure cascades in power networks [1], [2]. Current industry practice for failure prevention and mitigation mostly relies on $N - 1$ security constrained OPF as well as simulation-based contingency analysis [3], which are often implemented in tertiary control and operates on a slow timescale. As a result, such control usually prescribes a conservative operation point that strictly prohibits line overloads in all possible scenarios.

Since tertiary control can take effect 5 minutes to more than an hour after the disturbance, most literature on cascading failure analysis adopts the assumption that injections remain unchanged after a line failure if the post contingency network remains connected (and are changed only after a bridge failure according to a generic balancing rule that re-balances power in each island) [1], [4]–[6]. This assumption, however, is unrealistic and tends to be pessimistic: it does not take into account of frequency control mechanisms that adjust injections of controllable generators and loads immediately in response to line outages, on a faster timescale than that of post-contingency line tripping.

In this paper, we augment the existing steady state cascading failure models with frequency control dynamics that affect power flow redistribution in the new equilibrium post contingency. This integrated failure model is not only

more realistic, but also offers additional means to mitigate cascading failure through better design of the frequency control mechanism. Our proposed control strategy builds upon this extra freedom and reacts to line outages on a timescale of minutes.

Contributions of this paper: *We integrate a distributed frequency control strategy with a tree-partitioned network to provide provable failure mitigation and localization guarantees on line failures.* This strategy operates on a different timescale and supplements current practice, improving both grid reliability and operation efficiency. To the best of our knowledge, this is the first attempt to leverage results from the frequency regulation literature in the context of cascading failures, bringing new perspectives and insights to both literature. Our proposed strategy guarantees that (a) whenever it is feasible to avoid it, line failures do not propagate, and (b) the impact of line failures is localized as much as possible in a manner configurable by the system operator. A preliminary version of this work is presented in [7]. In this paper we include detailed proofs that were omitted in [7] and extend our simulations to further illustrate the performance of our approach.

We introduce the main idea of this new mitigation strategy in Section IV, which makes use of the so-called *Unified Controller (UC)*, a recent mechanism developed in the frequency regulation literature [8]–[12]. We specifically leverage the ability of UC to enforce line limits on a faster timescale than thermal line failure dynamics whenever possible. Our design revolves around the properties that emerge when the balancing areas that UC manages are connected in a tree structure. More specifically, in Section V, we characterize how UC responds to an initial failure, and prove that any *non-critical failure* is automatically mitigated and localized. Later, in Section VI, we discuss how the system operator can configure its mitigation strategy to minimize the impact of *critical failures*, and show 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 the fast-timescale of the frequency dynamics and the slow-timescale of the line tripping process. We further prove new results on the UC optimization problem with the spectral representation of DC power flow equations. Lastly, we apply classical results from convex optimization to show that critical failures can be detected in a distributed fashion.

In Section VII, we compare the proposed control strategy with classical Automatic Generation Control (AGC) using the IEEE 118-bus and 2736-bus test networks. We demonstrate that by switching off only a small number of transmission lines and adopting UC as the frequency 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 balancing area

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

where they occur, leaving the operating points of all other areas unchanged. This decoupling across balancing areas is important in practice. Lastly, we highlight that when load shedding is necessary, the proposed strategy incurs significantly smaller load loss.

Load shedding as a mitigation approach for cascading failures has been studied in the literature, e.g., using reinforcement learning [13], by shedding only the interruptible part of each load [14], by adaptively using affine control based on observed states [15], by formulating it as a DC OPF problem [16], or by treating the cascading process as a discrete-time optimal control problem [17]. In this paper we use the same DC power flow model as [15]–[17], but, unlike these works, we do not propose separate load shedding schemes to be implemented on a slow (power flow) timescale. Instead, load and generation control under the UC framework is performed as part of frequency regulation, during both normal operation and a contingency.

II. AN INTEGRATED FAILURE PROPAGATION MODEL

In this section, we present an integrated cascading failure model that incorporates frequency control dynamics and generalizes the steady-state DC failure model.

A. Fast-timescale Dynamics

Consider a power transmission network described by a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is the set of buses and $\mathcal{E} \subset \mathcal{N} \times \mathcal{N}$ are transmission lines. Using the notation in Table I, the post-contingency linearized frequency dynamics are:

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

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

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

The above differential equations model the fast-timescale response of the system to a transmission line failure. Note that we assume generator voltage control is local and converges at a faster timescale than generator-load frequency control. Therefore voltages are assumed to be fixed at their nominal values at the frequency control timescale.

The post-contingency injection deviation $p_j(t) := r_j + d_j(t)$ is the sum of the post-contingency disruption r_j and the system response $d_j(t)$. The disruption $r := (r_j, j \in \mathcal{N})$ comes from the effect of transmission line failures. In particular, suppose a failure of line (s,t) with pre-contingency flow f_{st}^{pre} happens, the post-contingency disruption is $r_s = f_{st}^{\text{pre}}$, $r_t = -f_{st}^{\text{pre}}$ and $r_j = 0$ for all other entries. The vector $d(t) := (d_j(t), j \in \mathcal{N})$ models frequency control and their values are determined by a feedback control mechanism (a non-controllable constant-power load is simply a special case where the controls are set as $d_j(t) \equiv d_j$). 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 can be feasibly mitigated (see Section VI for more discussion).

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^*.

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/controllable load deviations from pre-contingency values for generator/load buses
$p := (p_j, j \in \mathcal{N})$	aggregate post-contingency injection deviation
$\bar{d}_j, \underline{d}_j, j \in \mathcal{N}$	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
$\bar{f}_e, \underline{f}_e, e \in \mathcal{E}$	upper and lower limits for branch flow deviations
$n := |\mathcal{N}|$	number of buses
$m := |\mathcal{E}|$	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
$B := \text{diag}(B_e, e \in \mathcal{E})$	branch flow linearization coefficients that depend on line susceptances, nominal voltage magnitudes and reference phase angles

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

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

In other words, x^* satisfies the DC power flow model.[1]

The equilibrium to which the closed-loop system (1) converges thus determines the post-contingency DC power flow solution and can in turn impact how failures propagate in the network. A key insight from [8]–[12] is that the closed-loop equilibrium x^* of (1) is also an optimal solution of a certain DC-based 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 remark that the model (1) can be extended to include load buses j where $M_j = 0$. This system of differential algebraic equations assumes that the load dynamics evolve at a faster timescale than generator dynamics and are described by power flow equations. Including load buses does not change the stability properties of the closed-loop system, though the argument is slightly more complicated [8], [11].

B. Failure Propagation

In full generality, the failure model and the control strategy we introduce later apply to both generator failures and line failures, but, to simplify the presentation, in this paper we focus only on the latter. We describe the cascading failure process by keeping track of the set of outaged lines $\mathcal{B}(\mathfrak{n}) \subset \mathcal{E}$

---

[1]In primary frequency control literature (see [8], [10] 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 I for more details.

over stages $\mathfrak{n} \in \{1, 2, \ldots, N\}$ at steady state. Following a line outage, we assume that the system evolves on a fast timescale according to the frequency dynamics (1) during the transient phase. When it eventually converges to a closed-loop equilibrium, overloaded lines are tripped and the cycle repeats, as illustrated in Fig. 1.

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 $\mathcal{G}(\mathfrak{n}) = (\mathcal{N}, \mathcal{E} \setminus \mathcal{B}(\mathfrak{n}))$. If at equilibrium all branch flows $f^*(\mathfrak{n})$ are within the corresponding line limits, 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 at steady state, so we assume they trip at the end of stage $\mathfrak{n}$ and set $\mathcal{B}(\mathfrak{n}+1) = \mathcal{B}(\mathfrak{n}) \cup \mathcal{F}(\mathfrak{n})$. 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 [10] (spanning only seconds to a few minutes). This process then repeats for the subsequent stages. We remark that our model captures a simplified line trip dynamic where, after an initiating transmission failure (consisting of one or more line trips), the next set of (one or more) line trips does not occur until the frequency control dynamics has converged. In particular our model does not capture the finer thermal dynamics that determines the timing of line trips based on line currents and ambient temperatures.

The crux of our failure propagation model lies in the interplay between the slow-timescale line tripping process and the fast-timescale dynamics. 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 processes. For instance, as shown in Appendix I, if we adopt droop control for $d(t)$, the failure model in [6], [18] (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 non-cut failure. Since traditional AGC does not enforce line limits (congestion is managed on 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 on a fast timescale. 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 V that it will in fact localize the impact even when the outaged lines disconnect the network.

III. THE BRIDGE-BLOCK DECOMPOSITION AND THE UNIFIED CONTROLLER

The bridge-block decomposition and the unified controller have recently emerged as two important tools for grid reliability [10], [19], [20]. The two concepts operate on different timescales to improve the power system robustness:
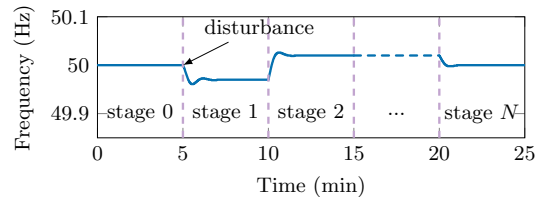

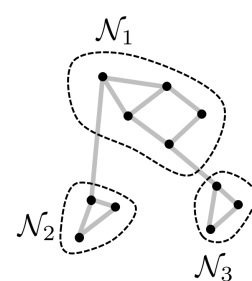Fig. 1: An illustration of the failure propagation model.

Fig. 2: Bridge-block decomposition of a graph.

the bridge-block decomposition aims to localize the failure propagation, while the unified controller aims to stabilize a disturbed system. In this section, we review these concepts and elaborate on how they can be integrated as a novel control framework for failure localization and mitigation.

A. Bridge-block Decomposition

Given a power network $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, a *partition* of $\mathcal{G}$ is defined as a finite collection $\mathcal{P} = \{\mathcal{N}_1, \mathcal{N}_2, \cdots, \mathcal{N}_k\}$ of nonempty and disjoint subsets of $\mathcal{N}$ such that $\bigcup_{i=1}^k \mathcal{N}_i = \mathcal{N}$. For a partition $\mathcal{P}$, each edge can be classified as either a *tie-line* if the two endpoints belong to different subsets of $\mathcal{N}$ or an *internal line* otherwise.

We define an equivalence relation on $\mathcal{N}$ such that two nodes are in the same equivalence class if and only if there are two edge-disjoint paths connecting them. For this specific partition, the tie-lines connecting different components are exactly the *bridges* (cut-edges) of the graph. We thus refer to this partition as *bridge-block decomposition* $\mathcal{P}^{\mathrm{BB}}$ of the power network (see Fig. 2 for an example).

It is shown in [19] that each graph has a unique bridge-block decomposition, which can be found in linear time. In particular, the bridge-block decomposition encodes rich information on failure propagation.

B. Unified Controller (UC)

UC is a control approach recently proposed in the frequency regulation literature [8]–[12]. Compared to classical droop control or AGC [21], UC simultaneously integrates primary control, secondary control, and congestion management on a fast timescale. The key feature of UC is that the *closed-loop* equilibrium of (1) under UC solves the following optimization

problem 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 [21], (2f) and (2g) are the flow and control limits. The matrix E encodes balancing area information as follows. Given a partition $\mathcal{P}^{\mathrm{UC}} = \{\mathcal{N}_1, \mathcal{N}_2, \cdots, \mathcal{N}_k\}$ of $\mathcal{G}$ that specifies the balancing areas for secondary frequency control, $E \in \{0,1\}^{|\mathcal{P}^{\mathrm{UC}}| \times n}$ is defined by $E_{lj} = 1$ if bus j is in balancing area $\mathcal{N}_l$ and $E_{lj} = 0$ otherwise. 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 [8]–[12] 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 (cut or non-cut) failure, the post-contingency system is driven by UC to an optimal solution of (2) (under appropriate assumptions). We refer the readers to [8]–[12] for specific controller designs and their analysis.

C. Connecting UC and the Bridge-Block Decomposition

We have introduced two partitions of a power network: the bridge-block decomposition $\mathcal{P}^{\mathrm{BB}}$ and the balancing area partition $\mathcal{P}^{\mathrm{UC}}$, which in general are different from each other. However, when they do coincide, the underlying power grid inherits analytical properties from both bridge-block decomposition and UC, making the system particularly robust against failures. Our proposed control strategy leverages precisely this feature, as we present in Section IV.

In practice, the balancing areas over which UC operates are usually connected by multiple tie-lines in a mesh structure. However, in order to align with the bridge-block decomposition, we may have to switch off a few tie-lines of $\mathcal{P}^{\mathrm{UC}}$. 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 [22] for more details. We henceforth assume that $\mathcal{P}^{\mathrm{BB}} = \mathcal{P}^{\mathrm{UC}}$. We refer to such a network as the *tree-partitioned* network since the balancing areas are connected in a tree structure prescribed by its bridge-block decomposition.

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 an area being "local" with respect to an initial failure.

Definition 3. *Given an initial failure $\mathcal{B}(1)$, we say that a tree-partitioned balancing area $\mathcal{N}_l$ 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 areas are adjusted whenever possible.

IV. PROPOSED CONTROL STRATEGY: SUMMARY

Our strategy consists of two phases: a planning phase in which tree-partitioned networks are created and an operation phase during which UC actively monitors and autonomously reacts to line failures during its operation.

A. Planning Phase: Align Bridge-blocks and 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. This is usually achieved via the zero area control error constraint in secondary frequency control [21], which is enforced by UC with (2e). As mentioned in previous sections, such balancing areas typically do not form bridge-blocks, as redundant lines are believed to be critical in maintaining $N-1$ security of grid [1], [2], [21].

We propose to create bridge-blocks whose components coincide with the balancing areas over which UC operates. This can be done by switching off a small subset of the tie-lines so that areas are connected in a tree structure. 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 bridge-blocks already in place.[2] It is interesting to notice that when the subset of lines to switch off is chosen carefully, the tree-partitioned network 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 VII-A.

We remark that it is natural to match bridge-blocks to balancing areas in transmission systems. First, there are typically small number of tie-lines between areas. Switching off those tie-lines is less likely to cause severe congestion to the grid. Second, each balancing area is supposed to balance its own power and the primary goal of failure mitigation is to prevent, as much as possible, a failure from impact other areas. This reconciles with the area control error (ACE) in traditional secondary frequency control.

In practice, the switching actions should be performed judiciously without jeopardizing the network operations. Any line switching action leads to power redistribution on remaining lines and may cause overload. It is thus crucial to assess the impact of each set of switching actions and avoid creating any congested line. It should be noted that switching

---

[2]In fact, line failures can lead to a "finer" bridge-block decomposition, as more bridge blocks are potentially created when lines are removed from service.

off lines may not necessarily increase the transmission cost [23]. Moreover, the system congestion level may even decrease when carefully choosing tie-lines to switch off [22]. We refer interested readers to [22] for an efficient algorithm to optimize such a line selection.

B. Operating Phase: Extending Unified Controller

Once the tree-partitioned areas are formed, the power network operates under UC as a closed-loop system and responds to disturbances such as line failures or loss of generator/load in an autonomous manner. Unlike traditional secondary frequency control, UC is distributed and reacts continuously to any disturbances as part of normal operation. There is no explicit termination of UC primal-dual dynamics. We assume that UC can detect line failures in real-time and thus react to the post-contingency network accordingly. In normal conditions where the system disturbances caused by line failures 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 successfully mitigated.

However, in extreme scenarios where a major disturbance (e.g., a critical failure) happens, the optimization problem (2) 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 8) and may lead to successive failures. There is therefore a need to extend the version of UC proposed in [8]–[12] 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 VI-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 areas to be progressively involved and coordinated in a systematic fashion when mitigating critical failures. We discuss some potential schemes in Section VI-B.

C. Guaranteed Mitigation and Localization

We show in detail in Sections V and VI that the proposed strategy provides strong guarantees in the mitigation and localization of both non-critical and critical failures. Specifically, it ensures that the cascading process is always stopped (a) after a non-critical failure by the associated areas, and the operating points of non-associated areas are not impacted in equilibrium or (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.

We remark that our strategy can maintain the $N-1$ security standard even though the balancing areas are connected in a tree structure. Unlike other classical approaches where failures that disconnect the network tend to incur more severe impact, our strategy can mitigate such failures as much as possible by autonomously adjusting the injections to rebalance and stabilize the system in each of the surviving component. In

fact, our fast-timescale control can significantly improve the grid reliability in $N-k$ sense as we show in the numerical experiments in Section VII.

V. LOCALIZING NON-CRITICAL FAILURES

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

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.

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 of the network.*

Proof. 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 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 an area $\mathcal{N}_v$ from the partition $\mathcal{P}$, 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 a tree-partitioned network that any area $\mathcal{N}_v$ from $\mathcal{P}$ is either within $\mathcal{C}_1$ or within $\mathcal{C}_2$, and that 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 subset of areas 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 $(j,i) \in \mathcal{E}$.

Note that (2e) ensures the injections are balanced for all areas. We can thus sum over all areas within $\mathcal{C}_1$:

$$0 = \mathbf{1}_{\mathcal{P}'}^T E C f^*$$

Following the definition of matrices E, C and the above notations, we can rewrite the above equations as:

$$
\begin{aligned}
0 &= \sum_{l:\mathcal{N}_l \in \mathcal{P}'} \sum_{i \in \mathcal{N}_l} \Big(\sum_{j:j \to i} f_{ji}^* - \sum_{j:i \to j} f_{ij}^* \Big) \\
&= \sum_{i:i \in \mathcal{C}_1} \Big(\sum_{j:j \to i} f_{ji}^* - \sum_{j:i \to j} f_{ij}^* \Big) \quad (3)
\end{aligned}
$$

Mathematically, $\left(\sum_{j:j \to i} f_{ji}^* - \sum_{j:i \to j} f_{ij}^* \right) + r_i + d_i^* = 0$ is the flow conservation for node i. Therefore, the right hand side of equation (3) calculates the summation of injections over all nodes within the connected component $\mathcal{C}_1$, which should be zero enforced by (2e).

Now, consider the node j_1 and the bridge $e = (j_1, j_2)$. For all other transmission lines (i,j) within component $\mathcal{C}_1$, the flow will be counted twice (for nodes i and j respectively) with opposite direction in (3) and thus gets cancelled out. Only the flow of bridge e is counted once for node j_1. Therefore, we conclude that the flow on the bridge must be zero, i.e. $f_e^* = 0$. Since the bridge e is arbitrary, we have thus proved the desired result. □

This lemma shows that tree-partitioned areas enable UC to achieve more than what it was originally designed for in [8]–[12]: the extended UC not only enforces zero area control errors through (2e), it also guarantees zero flow deviations on all bridges.

The following proposition is another result of this type, which clarifies how the tree-partitioned network induces a localization property under UC.

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 an area $\mathcal{N}_l$ is not associated with $\mathcal{B}(1)$, then at equilibrium $x^*(1)$ we have $d_j^*(1) = 0$ for all $j \in \mathcal{N}_l$.*

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

First, we construct a different point $\tilde{x}^* = (\tilde{\theta}^*, \tilde{\omega}^*, \tilde{d}^*, \tilde{f}^*)$ by changing certain entries of x^* within a non-associated area $\mathcal{N}_l$ 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 to show $\tilde{f}^* = BC^T\tilde{\theta}^*$. To do so, we first establish the following lemma:

Lemma 6. *For any tree-partitioned area $\mathcal{N}_z$ 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. If $\mathcal{N}_z = \mathcal{N}_l$, we have $\tilde{p}_j^* = 0$ for $j \in \mathcal{N}_l$ by construction and hence $\sum_{j \in \mathcal{N}_z} \tilde{p}_j^* = 0$. For $\mathcal{N}_z \neq \mathcal{N}_l$, 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. $\square$

Consider now an area $\mathcal{N}_w$ 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 6, 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 $CBC^T(\theta^* - \tilde{\theta}^*) = p^* - \tilde{p}^*$.

Lemma 7. *Let $\mathcal{P} = \{\mathcal{N}_1, \mathcal{N}_2, \cdots, \mathcal{N}_k\}$ be the tree-partitioned areas of $\mathcal{G}$ and consider a vector $p \in \mathbb{R}^n$ such that $p_j = 0$ for all $j \in \mathcal{N}_1$ and $\sum_{j \in \mathcal{N}_z} p_j = 0$ for $z \neq 1$. Then the Laplacian equation $CBC^T\theta = p$ is solvable, and any solution θ satisfies $\theta_i = \theta_j$ for all $i, j \in \overline{\mathcal{N}}_1$, where $\overline{\mathcal{N}}_1 := \{j : \exists i \in \mathcal{N}_1 \text{ s.t. } (i,j) \in \mathcal{E} \text{ or } (j,i) \in \mathcal{E}\}$.*

Proof. It is well-known (see [24] for instance) that the Laplacian matrix $L := CBC^T$ of a connected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ has rank $|\mathcal{N}| - 1$, and $L\theta = p$ is solvable if and only if $\mathbf{1}^T p = 0$, where $\mathbf{1}$ is the vector with proper dimension that consists of ones. Moreover, the kernel of L is given by $\text{span}(\mathbf{1})$.

If $\mathcal{N}_1$ is the only area in $\mathcal{P}$, then $p = 0$ since $p_j = 0$ for all $j \in \mathcal{N}_1$. We thus know the solution space to $L\theta = p$ is exactly the kernel of L, and the desired result holds.

If $\mathcal{N}_1$ is not the only area in $\mathcal{P}$, we can show that $f_e = 0$ for every bridge e connecting different areas. To see this, we start from a leaf area $\mathcal{N}_z$ of the tree-partitioned network. Since $\sum_{j \in \mathcal{N}_z} p_j = 0$, we have that the only bridge connecting the leaf area will carry zero flow. Therefore, the area $\mathcal{N}_z$ is decoupled from other areas and the Laplacian equation can be decomposed. We can iteratively use the above statement and decompose the Laplacian equation with respect to each area. With previous result, we know $\theta_i = \theta_j$ for all $i, j \in \mathcal{N}_1$ since $p_j = 0$ for all $j \in \mathcal{N}_1$. Together with the fact that bridges carry zero flow, we have $\theta_i = \theta_j$ for all $i, j \in \overline{\mathcal{N}}_1$. $\square$

By Lemma 7, we then have $\theta_j^* - \tilde{\theta}_j^*$ is a constant over $\overline{\mathcal{N}}_l$, and thus $\tilde{\theta}_i^* - \tilde{\theta}_j^* = \theta_i^* - \theta_j^*$ for all $i, j \in \overline{\mathcal{N}}_l$. 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 area $\mathcal{N}_l$. We have $\tilde{p}_j^* = 0$ by construction. From Lemma 6 we have $\sum_{j \in \mathcal{N}_z} \tilde{p}_j^* = 0$ for all z. Since $CBC^T\tilde{\theta}^* = \tilde{p}^*$, 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 $\tilde{f}_e^* = 0 = B_e(\tilde{\theta}_i^* - \tilde{\theta}_j^*)$. and therefore $\tilde{f}_e^* = B_e(\tilde{\theta}_i^* - \tilde{\theta}_j^*)$ for all $e \in \mathcal{E}$, concluding the proof. $\square$

Proposition 5 reveals that with the proposed control strategy, after a non-critical failure, the injections and power flows in non-associated areas remain unchanged at equilibrium, 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.

Furthermore, traditional control strategies usually treat failures that disconnects the system differently, i.e. the post-contingency injections normally stay constant, but can be changed if the system are disconnected due to the line failures [6]. Under our scheme, failures that disconnect the system are treated in exactly the same way as failures that do not, provided that they are non-critical. Moreover, the impact of a failure that disconnects the system is localized and properly mitigated to the associated areas as well. This is in stark contrast with the global and severe impact of a bridge failure [20] and is the key benefit of integrating UC with the bridge-block decomposition.

VI. 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 [8]–[12]. 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 [8]–[12], 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 $i \in \{1, 2, \cdots, n + 3m + |\mathcal{P}^{\mathrm{UC}}|\}$, be the dual variables corresponding to the constraints (2c)-(2f).

UC1: For all $j \in \mathcal{N}$, $\underline{d}_j \leq d_j(t) \leq \overline{d}_j$ 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).

Proposition 8. *Assume UC1 and UC2 hold. If* (2) *is infeasible, then there exists a dual variable* λ_i *such that* $\limsup_{t \to \infty} |\lambda_i(t)| = \infty$.

Proof. 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

$$\min_{\underline{d} \leq d \leq \overline{d}} \quad c(d) \quad \text{s.t.} \quad Ax \leq g, \ Cx = h \quad (4)$$

where A, C, g, h are matrices (vectors) of proper dimensions that can be recovered from the full formulation in (2). Let λ_1, λ_2 be the corresponding dual variables, 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

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

By the assumption UC2, we know that:

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

with the projection operator $[\cdot]^+_a$ defined component-wise by $([x]^+_a)_i = x_i$ if $x_i > 0$ or $(a)_i > 0$, and $([x]^+_a)_i = 0$ otherwise. Consider two closed convex sets $S_1 = \{x \mid Ax \leq g, Cx = h\}$ and $S_2 = \{x \mid \underline{d} \leq d \leq \overline{d}\}$. If the optimization (2) is infeasible, then $S_1 \cap S_2 = \emptyset$. As a result, we can find a hyper-plane that separates S_1 and S_2: more specifically, there exists $q \in \mathbb{R}^{3n+m}, q_0 \in \mathbb{R}$ such that

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

This fact then implies that the system $\{Ax \leq g, \quad Cx = h, \quad q^T x \leq q_0\}$ is not solvable. By Farkas' Lemma, we can thus find vectors w_1, w_2, w_3 such that $w_1 \geq 0, w_3 \geq 0$ (the inequality is component-wise), $A^T w_1 + C^T w_2 + qw_3 = 0$, 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:

$$\begin{aligned} z^T \dot{\lambda}(t) &= w_1^T [Ax(t) - g]^+_\lambda + w_2^T(Cx(t) - h) \\ &\geq w_1^T [Ax(t) - g]^+_\lambda + w_2^T(Cx(t) - h) \\ &\quad + w_3(q^T x(t) - q_0) \quad (5a) \\ &\geq w_1^T(Ax(t) - g) + w_2^T(Cx(t) - h) \\ &\quad + w_3(q^T x(t) - q_0) \quad (5b) \\ &= (A^T w_1 + C^T w_2 + qw_3)x(t) \\ &\quad - (w_1^T g + w_2^T h + w_3 q_0) = 0 + \epsilon > 0 \end{aligned}$$

where (5a) follows from $w_3 \geq 0$ and assumption UC1, which ensures $x(t) \in S_2$ and thus $q^T x(t) - q_0 \leq 0$, and (5b) comes from $w_1 \geq 0$ and the fact that $[x]^+_\lambda \geq x$ for all x (the inequality is component-wise). Consequently, $z^T \dot{\lambda}(t) - z^T \lambda(0) > \epsilon t$ and thus $\lim_{t \to \infty} z^T \lambda(t) = \infty$. Finally, by noting $\lim_{t \to \infty} z^T \lambda(t) \leq w_1^T \limsup_{t \to \infty} |\lambda_1(t)| + |w_2|^T \limsup_{t \to \infty} |\lambda_2(t)|$, the desired result follows. $\square$

Proposition 8 implies that, after a critical failure, UC cannot drive the system to a proper and safe operating point. This type of instability suggests a way to detect critical failures. Specifically, since Proposition 8 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 is exceeded. By doing so, critical failures can be detected in a distributed fashion during the normal operations of UC.

Note that the dual variables for non-critical failures are usually bounded in practice unless certain degeneracy in (2) happens. This, however, does not impact the stability of UC. That being said, non-critical failures may also cause relatively large dual variable values in transient states. The choice of detection thresholds thus inevitably involves trade-offs. Specifically, tighter thresholds allow critical failures to be detected more promptly so that the system only experiences a short period of instability. On the other hand, tighter thresholds also lead to a larger false alarm rate for non-critical failures. As we will show in the next section, when a failure is detected being critical, certain protection mechanisms will be involved so that the system is stabilized with potential cost of non-local response and load loss. Therefore, 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. 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 these two types of constraints, we 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

---

[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.
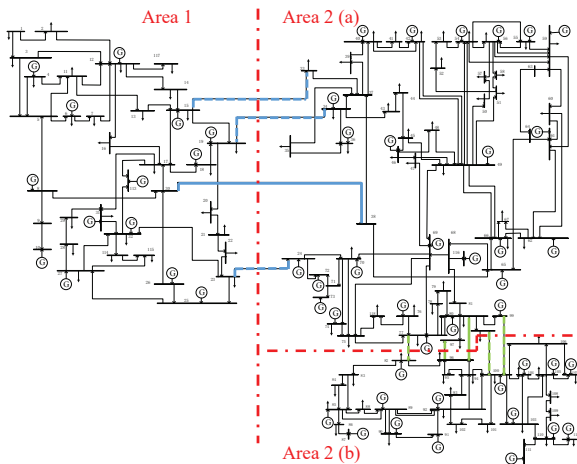
Fig. 3: One line diagram of the IEEE 118-bus test network.

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 areas to meet demand and shed load only 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.

In practice, these two types of constraint lifting can be combined and implemented iteratively, following the predetermined rules from the system operator. Specifically, the system operator may assess the node importance of the grid and design a hierarchical constraint lifting procedure, which specifies the order for involving balancing areas and the allowance for injection adjustment. Once the procedure is specified, the real-time response of UC can be implemented accordingly by changing the dynamics of dual updates; see [11] for the detailed design of UC dynamics.

VII. 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 Area 1 and Area 2 in Fig. 3. To form a tree-partitioned network, three lines $(15, 33)$, $(19, 34)$, and $(23, 24)$ are switched off, obtaining what we henceforth refer to as the *revised* network.

We compare UC on the tree-partitioned revised network, as specified by our proposal, 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. Note that we do not perform time-domain simulations, but directly compute the equilibrium point of the closed-loop systems under UC and AGC by solving the coressponding optimization respectively. This reduces the computational complexity and allows us to systematically evaluate our proposed control method in a wide range of simulation settings. Moreover, we assume that the detection threshold is set relatively large so that the failures can be accurately classified into critical and non-critical failures. Therefore, constraint lifting only happens for critical failures.

The failure scenarios are created as follows. First, we generate a variety of load injections by adding random perturbations to the nominal load profile from [25] and then solve the DC OPF to obtain the corresponding generator operating points over the actual IEEE test networks. For a fair comparison, we use the same generation and load operating points over the revised network and calculate the pre-contingency flows with DC power flow equations. 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. Specifically, 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, our simulations cover the cases $k = 1, 2, 3$ with roughly 138,600 failure scenarios, as summarized in Table. II.

TABLE II: 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

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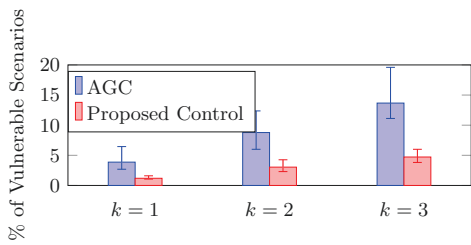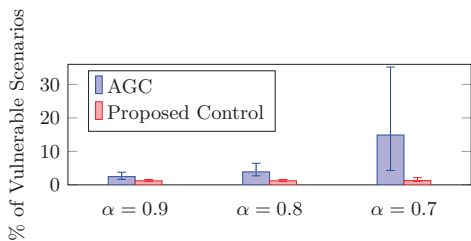
(a) Average fraction of vulnerable scenarios for $k = 1, 2, 3$.
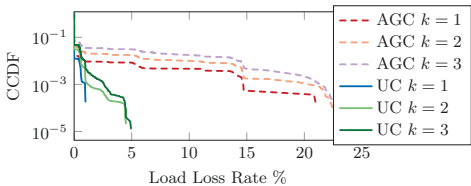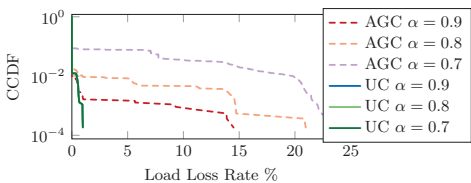

(b) CCDF for load loss rate for $k = 1, 2, 3$.

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.

outperforms AGC in all scenarios, especially those in which the system is congested. Again in all these scenarios, the bridge is never vulnerable under the proposed control.

We remark that UC outperforms AGC even if both control methods operate over the tree-partitioned network since AGC does not enforce line limits. There are only a few failure scenarios with a lower load loss rate under AGC. However, those failures always propagate through multiple stages and require a significantly longer time to re-stabilize the system.

B. Localized Failure Mitigation

In this subsection, we consider a specific constraint lifting rule that progressively involves other areas by relaxing area control error constraints only if local load shedding within the associated areas is not enough to make problem (2) feasible. This rule prioritizes localization of the initial failure. By implementing it, we show that the proposed control strategy can localize cascading failures within the associated areas with negligible load loss. The experiments are carried out
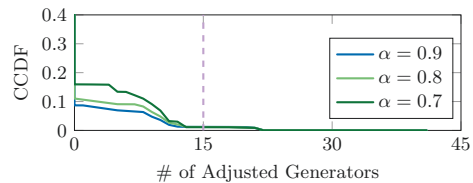
(a) CCDF for generator response on IEEE 118-bus network.
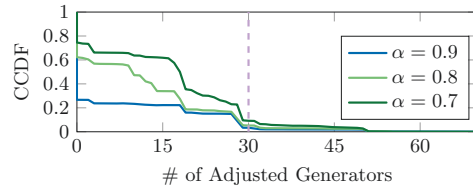
(b) CCDF for generator response on IEEE Polish network.

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

TABLE III: Statistics on failure localization over the 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)

over two networks: (a) a finer tree-partitioned version of the IEEE 118-bus test network, and (b) the much larger-scale Polish network, consisting of 2736 buses and 3504 lines.

For the IEEE 118-bus test network, we switch off 4 additional lines, $(77, 82)$, $(96, 97)$, $(98, 100)$, and $(99, 100)$, which refines the bridge-blocks used in the previous subsection since it further decomposes Area 2 into two balancing areas (as shown in Fig. 3). 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 load loss rates (LLRs) for this experiment are summarized in Table III. We observe that the proposed control strategy never incurs 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 areas, 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 on less than 15 generators, which is roughly the number of generators within a single balancing area. The small portion of failures that impact more than 15 generators are bridge failures, which by definition have two associated areas and thus more associated generators.

For the Polish network, we switch off 78 transmission lines from the original network, creating a tree-partitioned network with 4 areas 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

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

Line Capacity	$\alpha = 0.9$	$\alpha = 0.8$	$\alpha = 0.7$
Scenarios Mitigated with one Area (%)	92.39	88.63	86.91
Scenarios Mitigated with 2-3 Areas (%)	6.44	9.48	10.40
Scenarios Mitigated with All Areas (%)	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

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 IV. Our results show that for this test network, more than 86% of the single line failures can be mitigated locally within a single area 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.

VIII. CONCLUSION

In this paper, we have proposed a complementary approach to grid reliability by integrating the network bridge-block decomposition and the unified controller for frequency regulation to achieve fast-timescale failure control. It provides strong analytical guarantees of both the localization and mitigation of failures. 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.

APPENDIX I
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:

$$M_j \dot{\omega}_j = r_j + d_j - D_j \omega_j - \sum_{e \in \mathcal{E}} C_{je} f_e, \quad j \in \mathcal{N} \quad (6a)$$

$$f_{ij} = B_{ij}(\theta_i - \theta_j), \quad (i,j) \in \mathcal{E}. \quad (6b)$$

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 [21] as the dynamics for d_j's in (6), the cascading failure models from [20] and previous literature such as [6], [18] can be readily recovered. Indeed, as shown in [10], the *closed-loop* equilibrium of (6) under droop control is the unique[4] optimal

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

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

$$\min_{\theta, \omega, d, f} \quad \sum_{j \in \mathcal{N}} \frac{d_j^2}{2Z_j} + \frac{D_j w_j^2}{2} \quad (7a)$$

$$\text{s.t.} \quad r - d - D\omega = Cf \quad (7b)$$

$$f - BC^T \theta = 0 \quad (7c)$$

$$\underline{p}_j \leq r_j - d_j \leq \overline{p}_j, \quad j \in \mathcal{N}, \quad (7d)$$

where Z_j's are the generators' participation factors [21]. By plugging (7c) into (7b), it is easy to check that any feasible point $x = (\theta, \omega, d, f)$ of (7) satisfies $\sum_j r_j = \sum_j (d_j + D_j \omega_j)$. Cauchy-Schwarz inequality then implies that

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

$$\leq \sum_{j \in \mathcal{N}} \left(\frac{d_j^2}{2Z_j} + \frac{D_j \omega_j^2}{2} \right) \sum_{j \in \mathcal{N}} (2Z_j + 2D_j),$$

for which equality holds if and only if

$$d_j = \frac{Z_j}{\sum_j (Z_j + D_j)} \sum_j r_j, \quad \omega_j = \frac{\sum_j r_j}{\sum_j (Z_j + D_j)}. \quad (8)$$

Therefore, if the control limits (7d) are not active, (8) is always satisfied at the optimal point $x^* = (\theta^*, \omega^*, d^*, f^*)$.
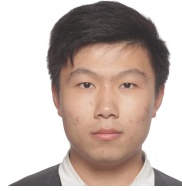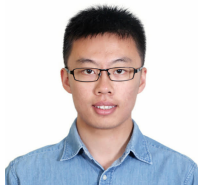
Now, consider a line e being tripped from the transmission network $\mathcal{G}$, and for simplicity assume the control limits (7d) are not active. If e is a bridge, the tripping of e results in two islands of $\mathcal{G}$, say $\mathcal{D}_1$ and $\mathcal{D}_2$, and two optimization problems (7) correspondingly. For $l = 1, 2$, $\sum_{j \in \mathcal{D}_l} r_j$ represents the total net power imbalance in $\mathcal{D}_l$, and therefore (8) 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 (8) 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 (7d) is active with a more involved analysis on the KKT conditions of (7). This droop control mechanism recovers the failure propagation model in [20] and underlies some of previous results in the literature on cascading failures in power systems [5], [6], [18].

REFERENCES

[1] 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.
[2] 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.
[3] R. Baldick *et al.*, "Initial review of methods for cascading failure analysis in electric power transmission systems ieee pes cams task force on understanding, prediction, mitigation and restoration of cascading failures," in *2008 IEEE PES General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*. IEEE, 2008, pp. 1–8.
[4] D. Bienstock and A. Verma, "The $n - k$ problem in power grids: New models, formulations, and numerical experiments," *SIAM Journal on Optimization*, vol. 20, no. 5, pp. 2352–2380, 2010.
[5] 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.
[6] S. Soltan, D. Mazauric, and G. Zussman, "Analysis of failures in power grids," *IEEE Trans. on Control of Network Systems*, no. 99, 2015.

[7] L. Guo, C. Liang, A. Zocca, S. H. Low, and A. Wierman, "Less is more: Real-time failure localization in power systems," in *IEEE Conference on Decision and Control (CDC)*, 2019, pp. 3871–3877.

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

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

[10] 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.

[11] E. Mallada, C. Zhao, and S. H. Low, "Optimal load-side control for frequency regulation in smart grids," *IEEE Trans. on Automatic Control*, vol. 62, no. 12, pp. 6294–6309, 2017.

[12] 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.

[13] 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, 2002.

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

[15] D. Bienstock, "Optimal control of cascading power grid failures," in *IEEE Conference on Decision and Control (CDC)*, 2011, pp. 2166–2173.

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

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

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

[19] L. Guo, C. Liang, A. Zocca, S. Low, and A. Wierman, "Line Failure Localization of Power Networks Part I: Non-cut outages," *IEEE Transactions on Power Systems*, 2021. [Online]. Available: https://doi.org/10.1109/tpwrs.2021.3066336

[20] L. Guo, C. Liang, A. Zocca, S. H. Low, and A. Wierman, "Line Failure Localization of Power Networks Part II: Cut Set Outages," *IEEE Transactions on Power Systems*, 2021. [Online]. Available: https://doi.org/10.1109/tpwrs.2021.3068048

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

[22] A. Zocca, L. Guo, C. Liang, A. Wierman, and S. H. Low, "A spectral representation of power systems with applications to adaptive grid partitioning and cascading failure localization," *arXiv preprint at arXiv:2105.05234*.

[23] E. B. Fisher, R. P. O'Neill, and M. C. Ferris, "Optimal transmission switching," *IEEE Trans. Power Systems*, vol. 23, no. 3, pp. 1346–1355, 2008.

[24] L. Guo, C. Liang, and S. H. Low, "Monotonicity properties and spectral characterization of power redistribution in cascading failures," in *Allerton Conference*, Oct 2017, pp. 918–925.

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

Linqi Guo received his B.Sc. in Mathematics and B.Eng. in Information Engineering from The Chinese University of Hong Kong in 2014, and his Ph.D. in Computing and Mathematical Sciences from California Institute of Technology in 2019. His research is on the control and optimization of networked systems, with focus on power system frequency regulation, cyber-physical network design, distributed load-side control, synthetic state estimation and cascading failure analysis.

Alessandro Zocca received his B.Sc. in mathematics from the University of Padua, Italy, in 2010, his M.A.St. in mathematics from the University of Cambridge, UK, in 2011, and his Ph.D. degree in mathematics from the University of Eindhoven, The Netherlands, in 2015. He then worked as postdoctoral researcher first at CWI Amsterdam (2016-2017) and then at California Institute of Technology (2017-2019), where he was supported by his personal NWO Rubicon grant. Since October 2019, he has a tenure-track assistant professor position in the Department of Mathematics at the Vrije Universiteit Amsterdam. His work lies mostly in the area of applied probability and optimization, but has deep ramifications in areas as diverse as operations research, graph theory, algorithm design, statistical physics, and control theory. His research focuses on dynamics and rare events on large-scale networked systems affected by uncertainty, drawing motivation from applications to power systems and wireless networks.
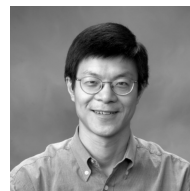
Steven H. Low (F'08) Steven Low is the F. J. Gilloon Professor of the Department of Computing & Mathematical Sciences and the Department of Electrical Engineering at Caltech and an Honorary Professor of the University of Melbourne, Australia. Before that, he was with AT&T Bell Laboratories, Murray Hill, NJ, and the University of Melbourne. He was a co-recipient of IEEE best paper awards, an awardee of the IEEE INFOCOM Achievement Award and the ACM SIGMETRICS Test of Time Award, and is a Fellow of IEEE, ACM, and CSEE. He was well-known for work on Internet congestion control and semidefinite relaxation of optimal power flow problems in smart grid. His research on networks has been accelerating more than 1TB of Internet traffic every second since 2014. His research on smart grid is providing large-scale cost effective electric vehicle charging to workplaces. He received his B.S. from Cornell and PhD from Berkeley, both in EE.

Chen Liang (S'19) received the B.E. degree in automation from Tsinghua University, Beijing, China, in 2016. He is currently pursuing the Ph.D. degree in Computing and Mathematical Sciences with the California Institute of Technology, Pasadena, CA, USA. His research interests include graph theory, mathematical optimization, control theory, and their applications in cascading failures of power systems.
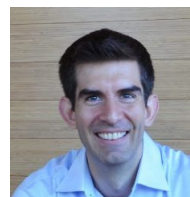
Adam Wierman is a Professor in the Department of Computing and Mathematical Sciences at the California Institute of Technology. He received his Ph.D., M.Sc. and B.Sc. in Computer Science from Carnegie Mellon University in 2007, 2004, and 2001, respectively, and has been a faculty at Caltech since 2007. He is a recipient of multiple awards, including the ACM SIGMETRICS Rising Star award, the IEEE Communications Society William R. Bennett Prize, and multiple teaching awards. He is a co-author of papers that have received best paper awards at a wide variety of conferences across computer science, power engineering, and operations research including ACM Sigmetrics, IEEE INFOCOM, IFIP Performance, and IEEE PES.