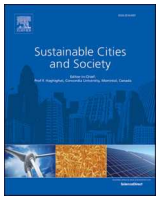
Sustainable Cities and Society 114 (2024) 105761

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs



Coordinated restoration of interdependent critical infrastructures: A novel distributed decision-making mechanism integrating optimization and reinforcement learning

Namrata Saha [a,1], Shabnam Rezapour [b,2,*], Nazli Ceren Sahin [b,3], M. Hadi Amini [a,1,4]

[a] Knight Foundation School of Computing and Information Sciences, Florida International University, Miami, FL, USA
[b] Enterprise and Logistics Engineering, Florida International University, Miami, FL, USA

ARTICLE INFO

Keywords:
Disaster management
Interdependent infrastructures
Resource allocation
Restoration scheduling
Reinforcement learning

ABSTRACT

The proper functioning of any society heavily depends on its critical infrastructures (CIs), such as power grids, road networks, and water and waste-water systems. These infrastructures consist of facilities spread across a community to provide essential services to its residents. Their spatial expansion and functional interdependencies make them highly vulnerable against natural/manmade disasters. Functional interdependencies mean that the functionality of components in one CI relies on the services provided by others. These features, combined with decentralized decision-making structure of CIs and the stochastic nature of post-disaster environments, highly complicate the optimization process for restoring CIs damaged in disasters. Optimizing CI restorations is critical to maximizing the post-disaster resilience of communities.

In this paper, we integrate and leverage Reinforcement Learning (RL) and optimization strengths to design a novel distributed modeling and solution approach for advancing the restoration process for interdependent CIs after disasters. The proposed approach (1) bridges the gap between integrative and distinct decision-making, enabling coordinated restoration planning for CIs within a decentralized decision-making context; (2) handles post-disaster uncertainties (e.g., uncertainty in recovery times of disrupted components); (3) generates adaptive solutions that cope with post-disaster dynamics (e.g., varying numbers of recovery teams); and (4) is flexible enough to handle several restoration decisions (e.g., restoration scheduling and resource allocation) simultaneously.

To evaluate its performance, we focus on restoring the road and power CIs in Sioux Falls, South Dakota, disrupted by several tornado scenarios. The numerical results show that coordinated policies in the restoration process of interdependent CIs consistently yield higher service for the community. The overperformance of the coordinated restoration policies can be as high as 27.9 %. The impact of coordination is more significant in severe disasters with higher disruptions and in the absence of efficient recovery resources.

1. Introduction

The proper functioning of any society heavily relies on its critical CIs, including road networks, power grids, and water/wastewater systems (Bush, 2003). Each CI comprises a set of physical components (e.g., cables, transmitters, and power generators in power CIs; roads, highways, bridges, and tunnels in road CIs; and pipes and water processing facilities in water/wastewater CIs) that span an area to provide key services to a community (Hafeznia & Stojadinović, 2023; Chertkov et al., 2015; Sharkey et al., 2015). Due to their spatial expansion and the increasing number of disruptive (natural or man-made) events, CIs are often subject to different types of disruption (Rezapour et al., 2021). For

* Corresponding author.
E-mail address: srezapou@fiu.edu (S. Rezapour).
[1] Postal address: FLORIDA INTERNATIONAL UNIV, KNIGHT FOUNDATION SCHOOL OF COMPUTING AND INFORMATION SCIENCES,11200 SW 8th Street CASE 238 Miami, FL 33199
[2] Postal address: FLORIDA INTERNATIONAL UNIV, COLL OF ENGR & COMPUTING, 10555 W FLAGLER ST # EC3114, MIAMI, FL 33174-1630
[3] Postal address: FLORIDA INTERNATIONAL UNIV, COLL OF ENGR & COMPUTING, 10555 W FLAGLER ST, MIAMI, FL 33174-1630
[4] Sustainability, Optimization, and Learning for InterDependent networks laboratory (solid lab)

https://doi.org/10.1016/j.scs.2024.105761
Received 13 April 2024; Received in revised form 16 July 2024; Accepted 18 August 2024
Available online 30 August 2024
2210-6707/Published by Elsevier Ltd.

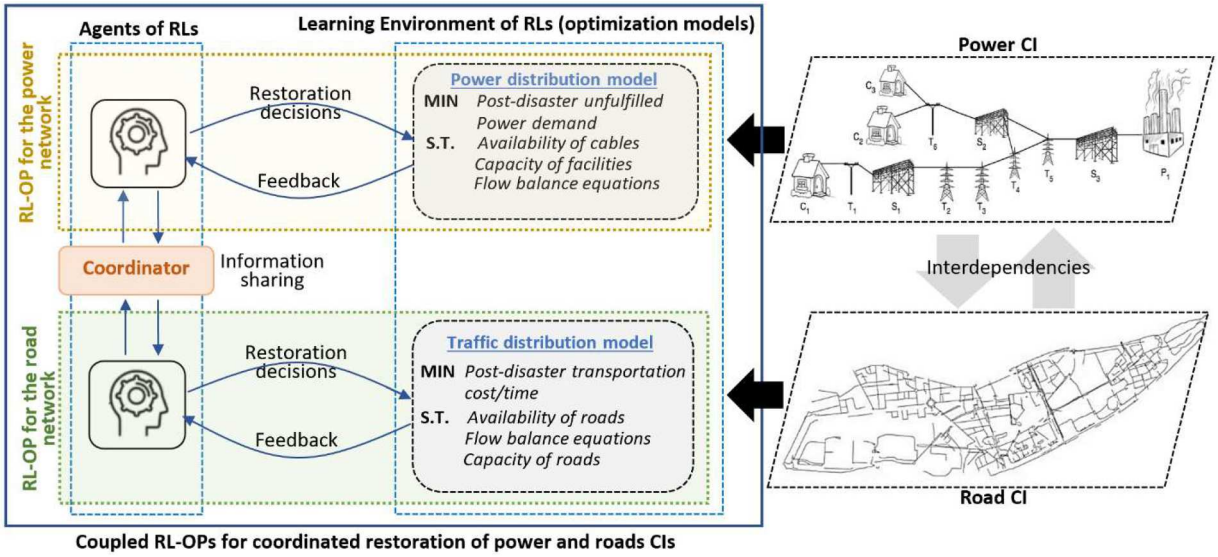


Fig. 1. The structure of the coupled RL-OPs for road and power CIs.

example, Hurricane Sandy, which hit the East Coast of the U.S. in 2012, significantly impacted several CIs: (1) the disruption in the power grid left 4.5 million customers without power for several days (Anon., Office of Electricity Delivery & Energy Reliability, 2012); (2) the closure of subway lines in New York City disrupted the road infrastructure (Kaufman et al., 2012); and (3) damage to wastewater treatment plans resulted in 10 billion gallons of sewage being spilled (Kenward et al., 2013). The total cost of restoring CIs damaged during this disaster was estimated at $65 billion (Anon., US Department of Housing & Urban Development, 2013). More recently, the tornado outbreak on March 31, 2023, affected at least eight states in the South and Midwest of the U.S., leaving 32 dead and dozens injured. Based on aggregated data from PowerOutage.us, >400,000 customers were without power. Although most roads were passable, traffic flow in the affected regions was very slow due to non-functional traffic signals (Moritz et al., 2023).

To enhance the resilience of communities against disruptive events, developing an efficient restoration policy for CIs is mandatory. The increasing interconnections among CIs have made them more vulnerable to disruptive events (Rinaldi et al., 2001; Fan et al., 2024), complicating their restoration efforts (Xu et al., 2024). CIs within a community are functionally interdependent, meaning the functionality of components in one CI depends on the services provided by other CIs (Huang & Wang, 2024). These interdependencies cause disruptions to cascade across communities (Amini et al., 2017; Sang et al., 2021; Lee II et al., 2007; Lee II et al., 2008; McDaniels et al., 2007, 2008). For example, a disruption in the power CI can lead to a wastewater treatment plant in the wastewater CI losing power and becoming inoperative, illustrating the propagation of disruptions across CIs. Similarly, restoration operations in CIs are interdependent. For instance, clearing roads of fallen trees in a road CI may only be possible after removing fallen power cables from the roads, a task carried out by the restoration crew of the power CI. This emphasizes the procedural interdependencies between restoration operations of CIs (Pinedo, 2012) and underscores the importance of coordinated restoration planning.

In practice, restoration efforts for CIs often disregard procedural interdependencies and are typically planned independently, with little to no communication (Leavitt & Kiefer, 2006; McGuire & Schneck, 2010). Recently, a few studies have focused on the concurrent restoration of interdependent CIs (Nurre et al., 2012; González et al., 2016; Cavdaroglu et al., 2013; Lee II et al., 2007; Garay-Sianca & Pinkley, 2021; Talebiyan & Dueñas-Osorio, 2023; Fan et al., 2024; Xu et al., 2024). Assuming a centralized decision-maker for all CIs, these studies primarily employ optimization techniques to schedule restoration efforts. However, the practical feasibility of this approach is very limited. CIs do not have full access to each other's information and conflicts of interest may arise among CIs due to differing priorities during restoration. For instance, a private power company might prefer to restore services to its higher-priority customers before addressing the power demands of other CIs. Observations from Emergency Operating Centers during previous disasters reveal that restoration decisions for CIs are made by different decision-makers, with local emergency managers facilitating communication among CIs (Sharkey et al., 2015). This highlights the necessity of coordinated restoration planning in a decentralized context for interdependent CIs. The importance of coordination and information-sharing in restoring CIs has already been underscored by many researchers (Caruson & MacManus, 2008; Somers & Svara, 2009; Kapucu & Garayev, 2013; Sharkey et al., 2015).

The restoration of CIs is conducted in challenging post-disaster environments. Due to resource scarcity (e.g., limited facilities, supplies, and manpower) following a disaster, the restoration of disrupted components cannot be initiated simultaneously (Oruc & Kara, 2018; Farzaneh et al., 2023; Aksu & Ozdamar, 2014; Sahin et al., 2016). Resource constraints typically lead to sequential restoration efforts that extend throughout the disaster response phase (Farzaneh et al., 2023). All restoration activities occur in chaotic post-disaster situations characterized by varying levels of uncertainty and dynamism, such as uncertain damage levels and restoration times for disrupted components, and the dynamic number of facilities and recovery crews available for restoration. These facts highlight the importance of developing modeling and solution techniques to address the necessity of generating sequential, stochastic, and adaptive restoration policies for CIs. These complexities are compounded in the coordinated restoration of interdependent CIs.

The above discussions demonstrate that coordinated restoration of interdependent CIs needs a distributed decision-making approach that preserves the autonomy and decentralization of CIs while enabling them to coordinate their decisions with limited information sharing. As will be discussed in detail in Section 2, the CI restoration literature lacks a decision-making approach to handle these requirements. To design such a distributed decision-making approach, the following research questions are answered in this paper:

- **Research question 1** - To address the problem of restoring interdependent CIs, what decision-making elements are needed within this distributed approach? To address the decentralization of CI restorations, this approach needs several distinct decision-making

elements, one for each CI. These elements should be able to make restoration decisions for their corresponding CI, operating in stochastic and dynamic post-disaster circumstances with limited restoration resources. As shown in Fig. 1, the decision-making element of each CI combines the strengths of Reinforcement Learning (RL) and optimization, called an RL-OP. This combination facilitates sequential, stochastic, and adaptive decision-making for each CI.

- **Research question 2** – To include interdependencies, how can we coordinate the restoration decisions made by distinct RL-OPs? To consider CI interdependencies in the decision-making process of each RL-OP, we will design and embed a "coordinator" within the distributed decision-making approach (see Fig. 1). The coordinator assists RL-OPs in partially exchanging restoration information, preventing them from making infeasible (or non-executable) decisions, and enabling the evaluation of the consequences of their decisions on other CIs. This results in a coupled RL-OPs that make coordinated decisions, improving the performance of all CIs in providing better services to the community, rather than distinct decisions only improving the performance of their own CIs.
- **Research question 3** – In comparison to current practices where CI interdependencies are ignored in the restoration processes, how much improvement is expected from the restoration policies generated by this novel distributed decision- making approach, coupled RL-OPs? The post-disaster performance of CIs in providing services to the community will be compared under two groups of distinct and coordinated restoration policies under several disruption scenarios. This comparison will provide a good estimation of the improvement achievable by using this new decision-making approach.

The application of the proposed approach is not limited to interdependent CIs and not just to restoration decisions. This modeling structure is generalizable for making coordinated decisions for a broad range of decentralized yet interdependent systems (DISs). A DIS represents a fusion of diverse, autonomous, yet interdependent systems operating in various physical, social, cyber, or technical contexts. Within this intricate arrangement, the behavior and performance of one system intricately intertwine with those of others. DISs find prominent applications in the contemporary world, including cyber-physical-social systems, the Internet of Things, and supply chains. The proposed mechanism can be widely used to optimize DIS decisions.

The paper's organization is as follows: Section 2 reviews the literature on CI restoration and highlights the paper's contributions. A detailed description of the problem is included in Section 3. The proposed coupled RL-OPs approach is developed in Section 4. Section 5 explains the case study, presents numerical results, and discusses derived insights. Section 6 concludes the research.

2. Literature review

The literature on resilient CIs comprises two groups of research: (1) pre-disaster preparation, which addresses strengthening and fortifying CIs to reduce their vulnerability against disasters (Dobson et al., 2001; Botterud et al., 2005; Chen et al., 2005; Bienstock & Mattia, 2007; Fang & Zio, 2019; Ouyang, 2017; Bhuiyan et al., 2020; Fakhry et al., 2022); and (2) post-disaster response, which pertains to the efficient restoration of damaged CIs. Of course, another possibility is the integration of pre- and post-disaster operations (Sütiçen et al., 2023). The problem investigated in this paper belongs to the post-disaster restoration of CIs. In this section, we review the papers of the second group and discuss the contributions of the research against two research streams in post-disaster CI restoration: (i) concurrent restoration of interdependent CIs (will be reviewed in Section 2.1) and (ii) optimization of CI restoration in stochastic and dynamic post-disaster circumstances (will be reviewed in Section 2.2).

2.1. Concurrent restoration of interdependent CIs

In this section, we review the decision-making context (e.g., centralized or decentralized) of models that have been proposed in the literature for concurrent restoration of interdependent CIs and the restoration decisions optimized by these models. There are some studies that only focus on modeling interdependencies and commodity flow (e.g., Lee II et al., 2007) or forecasting failure cascading (e.g., Loggins & Wallace, 2015) in interdependent CIs. They do not make any restoration decisions and are out of the scope of this paper. In this paper, the restoration decisions that deal with determining the sequence/-concurrency of recovering disrupted components are named "*restoration scheduling,*" and the decisions of assigning limited recovery resources (restoration crews, facilities, machineries, etc.) to the disrupted components selected to be recovered simultaneously are called "*resource allocation*".

Nurre et al. (2012) propose an optimization model to restore power, water, and emergency good supply networks concurrently. The model assigns disrupted links to a set of recovery crews to install them into the networks. The model pays no attention to the importance of links, limiting the assignment of only one team to each disrupted link for restoration. Resource allocation is not addressed in this model. In this model, restoration decisions for all CIs are centrally made by a single decision-maker with complete access to the information of all CIs. González et al. (2016) introduce the Interdependent Network Design and Scheduling (INDS) problem, focusing on designing a reconstruction strategy for a partially destroyed system of CIs. They propose a mixed integer optimization model to determine which disrupted components should be restored in each CI and the optimal sequence for restoring these components. This research does not address resource allocation decisions and makes all restoration decisions of CIs centrally. The authors solve the problem using a simulation-optimization approach, testing it on power, water, and gas CIs in Shelby County, TN, U.S.

Cavdaroglu et al. (2013) investigate the challenge of restoring power and telecommunication CIs following unexpected events that disrupt their services. They devise a mixed-integer optimization model to optimize the recovery sequence of disrupted links for each recovery group. The objective is to minimize the total costs associated with flow, unmet demand, and new installations throughout the restoration timeframe. In this model, each disrupted component is assigned to only one recovery team, and no decision is made regarding resource allocation. A single authority centrally makes restoration decisions for all CIs. Almoghathawi et al. (2019) focus on the restoration challenges in a network of interdependent power-water CIs following a disruptive event. They develop a multi-objective restoration model: the primary objective is to enhance the resilience of the CIs while minimizing the overall restoration costs. Again, they assume that each disrupted component can only be restored by a single team, without the possibility for multiple teams to collaborate simultaneously on components to expedite their recovery processes. The decision-making nature of the problem is centralized.

Garay-Sianca and Pinkley (2021) propose an integrated network design and scheduling problem for a system of two interdependent power and road CIs with the movement of restoration machines. They formulate the problem as mixed integer programming and solve it using a rolling horizon solution procedure. The model makes three key decisions for each CI: (i) identifying damaged links that should be restored, (ii) allocating the appropriate machinery for the restoration process of each link, and (iii) establishing the order of restoration tasks assigned to each machine. This model considers the dynamic movements of machinery as they navigate through the evolving layout of the road CI. The restoration decisions of CIs are made centrally by a single authority. Sütiçen et al. (2023) study pre-disaster reinforcement and post-disaster restoration within interdependent CIs. They model the problem as a scenario-based two-stage optimization model. In the initial stage, decisions regarding reinforcement are made, while in the subsequent stage, restoration activities are strategized for repair teams across

various potential disaster scenarios. Their approach assumes that simultaneous restoration of impaired links is prohibited. So, there is no need for resource allocation. Oversight of reinforcement and repair operations for interdependent CIs is entrusted to a central decision-maker possessing complete information.

Maraqa et al. (2022) introduce a multi-objective model to optimize the restoration sequence and crew allocation for interdependent CIs. However, this model empowers a singular governing authority possessing comprehensive information on all CIs to make informed decisions regarding their restoration process. Alkhaleel et al. (2022) develop a mixed-integer linear programming model to optimize the restoration process for a set of interdependent CIs, including power and water CIs. In this model, a single decision-maker with full CI information access identifies failed components for restoration and assign work crews to those components. The model allows multiple crews to work simultaneously on a single component to enhance system resilience. Fan et al. (2024) employ mixed-integer second-order cone programming to optimize the coordinated restoration process for interdependent power, gas, and transportation CIs. The decision-making structure of the model is centralized, with objectives to maximize the restored power and gas demand and enhance traffic load capacity in the community. The model only determines the sequence of restoring disrupted links in each CI. Huang and Wang (2024) address post-disaster centralized restoration planning for interdependent CI systems, specifically focusing on electric power and potable water networks. Using a bi-objective optimization model combined with Monte Carlo simulation, their research seeks Pareto-optimal solutions that balance trade-offs between losses in social services and economic production during the recovery process. The model determines the optimal sequence for restoring disrupted CI components. Xu et al. (2024) study centralized restoration planning for interdependent CI systems, focusing on electric power and gas systems. These systems are modeled as an undirected integrated network with bidirectional dependencies. Initially, a deterministic model is developed using mixed-integer linear programming to optimize the repair sequence and maximize system resilience. The model is then extended to a two-stage stochastic model that accounts for uncertainty in repair times, represented by a set of scenarios with known distribution functions.

In all the studies reviewed above, restoration decisions for CIs are made in a centralized context using optimization techniques by a single decision-maker with full access to the information of all CIs. Acknowledging the limitations of optimization, there are a few studies that employ other techniques to interdependent CI restoration. For example, Sun and Zhang (2020) propose a model that integrates agent-based simulation and RL to determine the optimal sequence for restoring disrupted components in a network of interdependent CIs, including transportation, power generation, and wastewater treatment facilities. Again, the proposed model operates in a centralized context, with a single RL agent making restoration decisions for all CIs. At each decision-making step, only one team can be assigned to each disrupted component. Hafeznia and Stojadinović (2023) propose the Resilience Quantification Iterative Optimization-based Simulation (ResQ-IOS) framework to study the seismic resilience of interdependent CI systems (CISs) in Shelby County, USA. This framework integrates simulation and optimization methods to assess resilience, emphasizing bi-directional interdependencies between power generation and natural gas production, and between power generation and water supply sectors. The simulation component employs a heuristic approach using a criticality-based strategy to specify the restoration sequence of damaged components. For example, nodes with the largest demand are repaired first and links with the largest capacity are repaired first. The optimization component only determines the optimal flow of resources and services from and to each node in CIs to minimize the loss of resilience at each step of the recovery process. The framework does not make any resource allocation decision. However, they just analyze the impact of changing resource availability in a few scenarios on the resilience of the CI system.

There are very few studies in literature considering decentralized decision-making for the restoration of interdependent CIs. One such study is Smith et al. (2020). They propose an ad hoc sequential game-theoretic model, representing a discrete time noncooperative game between CI decision-makers, to optimize restoration sequence in an interdependent CI system. In this formulation, restoration decisions of CIs are made sequentially by their corresponding decision-makers rather than concurrently. Additionally, the computational complexity of game models compels the authors to significantly simplify the restoration operations. For example, they assume that the required resources to recover all disrupted components (links or nodes) in all CIs are the same and equal to 1. Talebiyan and Dueñas-Osorio (2023) propose an auction-based approach to allocate recovery resources among a set of disrupted interdependent CIs in a decentralized fashion. Each decision-maker employs a mixed-integer optimization model to devise the minimum-cost restoration plan, considering resource and operational constraints. The objective function includes network flow cost, arc and node restoration cost, under and oversupply penalties, and site preparation cost. The auctions entail no communication among decentralized decision-makers, implying lack of coordination during the decision-making process. This is the main point differentiating their study from this paper.

This literature review highlights the lack of a systematic modeling and solution approach for coordinately restoring interdependent CIs in a decentralized context. We address this methodological gap by developing a new distributed decision-making approach capable of meeting the following requirements: (1) restoration decisions of CIs are made in a decentralized context (by separate decision-makers) but are coordinated through partial information sharing among CIs, (2) it preserves privacy of CIs by minimizing information sharing, and (3) the proposed approach is flexible enough to make several restoration decisions (e.g., restoration scheduling and resource allocation) simultaneously.

2.2. Optimization of CI restoration in stochastic and dynamic post-disaster circumstances

In this section, we review the post-disaster complexities (e.g., uncertainties and dynamics in restoration operations) considered in the literature in modeling CI restoration. We aim to highlight the additional contributions of the paper to the literature by discussing the extra flexibility added by the coupled RL-OPs approach to modeling restoration processes.

In the aftermath of a disaster, the environment is often chaotic, characterized by numerous uncertainties and dynamics arising from incomplete information and predictions. However, the majority of studies examining restoration operation, even in a single CI, operate under deterministic and static assumptions. For example, Averbakh (2012) and Averbakh and Pereira (2012) employ mixed-integer programming to optimize the recovery sequence for damaged links in a transportation CI by one and several recovery teams. The models are completely deterministic without any temporal variations. Matisziw et al. (2010) propose a deterministic multi-objective model to optimize the sequence of recovering damaged links and nodes in a CI. The model analyzes the tradeoff between two objective functions: recovery cost minimization and system flow maximization. They assume all nodes and links have the same recovery times and can be restored within a single time unit. Baxter et al. (2014) propose an integer programming model for the incremental reconstruction of a damaged network. The objective is to minimize the cost of clearing/opening damaged links and to minimize the penalty of not satisfying demands. The model is static and deterministic. Nurre and Sharkey (2014) investigate the problem of network-based CI restoration when several identical machines are working in parallel. This problem includes designing a network of facilities (each facility includes multiple machines) and efficiently scheduling restoration activities on machines. However, the number of facilities is fixed over the restoration horizon and the recovery times of

damaged components are known in advance.

Morshedlou et al. (2018) investigate the problem of routing recovery teams to recover disrupted links/nodes in a single CI. They develop two optimization models to dispatch and route recovery teams towards disrupted components in the CI to maximize network resilience progress over the restoration horizon. The number of recovery teams is fixed, and no uncertainty is included in the models. Fan et al. (2024) assume that the capacity of restoration resources in the CI restoration process is static and recovery time of disrupted components is deterministic and uniform. For example, at two-hour intervals per step, they assume that recovery teams in the power CI repair 2 links per step, those in the natural gas CI repair 1 link, and teams in the transportation CI recover 2 roads per step. Other recent studies, such as Sun and Zhang (2020) and Hafeznia and Stojadinović (2023), also assume the fixed recovery resources and fixed recovery times in their restoration planning models.

The number of studies considering post-disaster dynamics is very limited. For example, Aksu and Ozdamar (2014) propose a dynamic path-based model (formulated as integer programming) to maximize network accessibility in a road CI after a disaster. They assume that recovery times of disrupted components are different but deterministic. However, the number of recovery resources (e.g., teams and facilities) are dynamic and may change over time. Ulusan and Ergun (2018) present an innovative index inspired by network science to assess the criticality of components within a disrupted road CI. They propose a restoration heuristic aimed at prioritizing restoration activities according to this index. They ignore uncertainties and assume the presence of complete information about the debris level and recovery times of blocked links. However, the number of recovery teams may change over time.

Very little research addresses the challenge of incomplete information or uncertainty in disaster restoration operations. Xu et al. (2016) propose a stochastic integer model to schedule inspection, damage assessment, and repair tasks for optimizing post-earthquake restoration in a power CI. The objective is to minimize each customer's average time without power. The expected recovery time for each disrupted component is uncertain, and to simplify the model, they define several limited scenarios representing uncertainty in recovery times of disrupted nodes. Following a similar approach, Alkhaleel et al. (2022) define scenarios to represent uncertainty in the recovery times of disrupted components. Similarly, Huang and Wang (2024) employ Monte Carlo simulation to generate scenarios representing damage levels and repair times of disrupted components. Farzaneh et al. (2023) explore the challenges stemming from incomplete data on damage and the lack of coordination among post-disaster restoration operations. To address the lack of complete damage information, they employ a real-time damage assessment and data collection mechanism that requires pre-disaster Unmanned Aerial Vehicles (UAVs) prepositioning. Without a real-time damage assessment mechanism similar to our problem, considering uncertainty in damage levels and required recovery times significantly increases modeling accuracy. It avoids sub-optimality or even infeasibility of generated restoration policies when nominal values do not materialize for those parameters.

This literature review reveals the lack of a systematic modeling approach for CI restoration that can address complexities of post-disaster circumstances without imposing unrealistic simplifying assumptions (e.g., discrete scenarios for recovery times) and locating pre-disaster monitoring facilities (e.g., pre-positioned UAVs). The modeling and solution approach proposed in this paper can address these gaps and contribute to the literature on CI restoration in the following ways: (1) It is capable of handling post-disaster uncertainties in the recovery times of disrupted components; and (2) It generates adaptive solutions that cope with post-disaster dynamics of varying numbers of recovery teams.

2.3. Research contributions

The contributions of this paper to the CI restoration literature can be summarized as follows:

- This paper proposes the first systematic distributed decision-making approach, coupled RL-OPs, to generate coordinated restoration policies for a set of interdependent CIs operating in a decentralized context. This approach allows CI decision-makers to preserve the information privacy of their CIs while coordinating their policies with partial information sharing.
- The problem decomposition capability added to the proposed approach by RLs enables it to handle post-disaster complexities, such as uncertainty in the recovery time of disrupted components and the dynamic number of recovery teams over the restoration process of CIs.
- The proposed approach generates comprehensive restoration policies that not only determine the restoration schedule for disrupted components but also identify the best resource allocation to the components selected for concurrent restoration.

3. Problem description

Without sacrificing generality, in this section, we concentrate on developing coordinated restoration policies for interdependent road and power CIs (as displayed in Fig. 1). The structure of the disrupted power and road CIs, the restoration decisions for each CI, and their interdependencies are explained in the subsequent sections. However, it is worth noting that the mathematical and computational foundations of the proposed approach can be extended to any set of interdependent CIs managed within a decentralized context.

3.1. Power CI

We represent the power CI in the disaster-affected community using a network: $G^P(N^P, L^P)$. The set of nodes in the network includes supply nodes that generate power (N_S^P), intermediate nodes that transfer power (N_I^P), and demand nodes representing aggregated households in municipal sites (N_D^P): $N^P = N_S^P \cup N_I^P \cup N_D^P$. These nodes are connected through cables represented as links in the network, $l = (n, n')$ where n and $n' \in N^P$. The daily power generation capacity at each supply node is denoted by PC_n^P $(\forall n \in N_S^P)$. The parameter DD_n^P $(\forall n \in N_D^P)$ represents the daily demand at the demand nodes of the network. Additionally, there is a flow capacity for the links/cables of the network represented by TC_l^P $(\forall l \in L^P)$. Under power generation and transmission limitations, the power distribution plan in a power CI determines how the daily generated power at the supply nodes should be routed throughout the network to fulfill the daily demands materialized at the demand nodes.

Extreme events (e.g., thunderstorms, hail, lightning, tornados, and hurricanes) may damage the power CI by disrupting some of the links (e.g., downing some power lines) in the network. This may distort its power distribution plan and leave some of the demand nodes without power. A restoration plan for the power CI includes two groups of decisions: (1) restoration scheduling, which determines the sequence or concurrency of restoring disrupted links in the network, and (2) resource allocation, which determines the best scheme to assign recovery teams to the links selected for concurrent restoration. The objective of the restoration plan is to minimize the total unfulfilled power demand during the restoration period, represented by T. Limitation of restoration teams, uncertainty in the recovery times of disrupted links, interdependencies between the road and power CIs, and the decentralized decision-making structure of CIs are the key barriers complicating restoration optimization for power CIs.

The main assumption related to the power CI is that we employ a linear DC model to approximate the nonlinear AC model used for power distribution planning in the power network. The accuracy of this approximation has been shown by Bienstock and Mattia (2007).

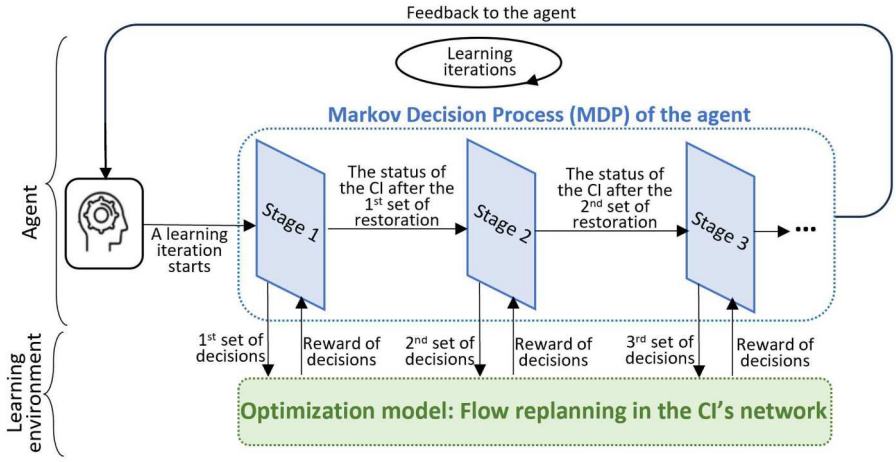


Fig. 2. The structure of the RL-OP for each CI.

3.2. Road CI

The road CI in the community is represented by another network, $G^R(N^R, L^R)$, in which urban sites constitute nodes of the network (N^R), connected through roads/highways represented by links (L^R). Each link has a traffic flow capacity ($FC_l^R,\ \forall l \in L^R$). When an extreme event damages the road CI, some roads/highways become disrupted and impassable, decelerating the post-disaster traffic flow in the affected area. The post-disaster traffic need is represented as a set of Origin-Destination (OD) pairs, where OD pairs correspond to daily traffic flow moving from origin (N_O^R) to destination (N_D^R) nodes of the road CI through intermediate nodes $(N_I^R) : OD = \{od = (m, m') | m \in N_O^R \text{ and } m' \in N_D^R\}$. The traffic demand of each pair od is denoted by $TF_{od}^R\ (\forall od \in OD)$.

Developing a restoration plan for a road CI involves strategizing the restoration schedule and resource allocation for disrupted roads to maximize the acceleration of post-disaster traffic flow within the restoration period, T. This optimization process must account for constraints such as limited and dynamic resources (e.g., a finite and dynamic number of recovery teams) and uncertainties such as incomplete information regarding the recovery time of blocked or damaged roads and highways.

In this paper, the focus is on short-term restoration of critical CIs that starts immediately after a disaster in a community. These operations should be accomplished within a short planning horizon after the disaster (e.g., within a couple of weeks) (FEMA, Anon., 2018). That is why we introduced the planning horizon T in the problem. Within this short interval after the disaster, the travel flow in the affected region is usually limited to relief operations such as transferring casualties to hospitals, transporting relief commodities from stocks and airports to affected regions, and relocating affected residents to shelters. Therefore, we assume that routine pre-disaster traffic flows that may cause traffic jams do not exist in the area. In Section 4.2.2, we will explain how this assumption can be relaxed in the problem.

3.3. Interdependencies between road and power CIs

We define several sets to model restoration interdependency between the road and power CIs. Π_l^P includes the set of prerequisite links in the road CI that should be restored before restoring link $l \in L^P$. Similarly, Π_l^R includes the set of prerequisite links in the power CI that should be restored before restoring link $l \in L^R$. In this paper, our goal is to coordinately optimize the restoration plans for the damaged power and road CIs, considering their interdependencies, to minimize the total unfulfilled power demand in the power CI and travel time/cost in the road CI. The restoration decisions for CIs are made in a decentralized context by separate decision-makers.

4. Problem formulation

Fig. 1 demonstrates the general structure of the approach developed to generate coordinated restoration policies for the road and power CIs, referred to as the coupled RL-OPs mechanism. This mechanism includes a distinct RL-OP for each CI dealing with its restoration decisions. Having a separate RL agent for each CI is consistent with the decentralized nature of the problem. Having a separate learning environment for each agent enables us to handle the heterogeneous operational environment of CIs. To harmonize the decisions of agents and generate coordinated policies, we will design and locate a "coordinator" to facilitate limited information sharing among the RL agents.

RL is a machine learning technique consisting of an agent and a learning environment. It trains the agent by using feedback from the learning environment to guide the agent toward optimal solutions. RL mimics the trial-and-error learning process employed by humans to achieve their goals (Li, 2023). For more information about the principles of RL refer to Ding et al. (2020), Meyn (2022), and Morales (2020).

Fig. 2 demonstrates the structure of the RL-OP that is generated for each CI in this paper. The agent of the RL-OP will make restoration decisions for the CI that include restoration schedule and resource allocation. Since scheduling is a sequential process (we want to identify the sequence of restoring damaged links), we model the decision-making structure of the agent of the RL-OP as a Markov Decision Process (MDP) which permits the agent to have several decision-making stages. In each stage, a new set of restoration decisions will be made by the agent for the CI. The consequence/reward of the restoration decisions made in each stage is evaluated in its learning environment. The learning environment is an optimization model that replans the flow (traffic or power flow) movement in the CI after restoring selected links and calculates the improvement in its performance (e.g., total reduction in the unfulfilled demand of the power CI or reduction in travel cost/time of the road CI). Fig. 2 shows the flow of information between the agent and the environment of an RL-OP in one learning iteration. These iterations train the agent to learn more about the reward achievable by making any decision in each stage of the RL-OP and gradually guide the agent to make better decisions with higher rewards in MDP stages. After a high number of iterations, the agent is trained enough to select the optimal decision for each MDP stage.

More details about the RL-OPs designed for the power and road CIs are respectively explained in Sections 4.1 and 4.2. Section 4.3 includes the development of the "coordinator" to harmonize the decisions of the RL-OPs' agents. Connecting agents of the RL-OPs through the

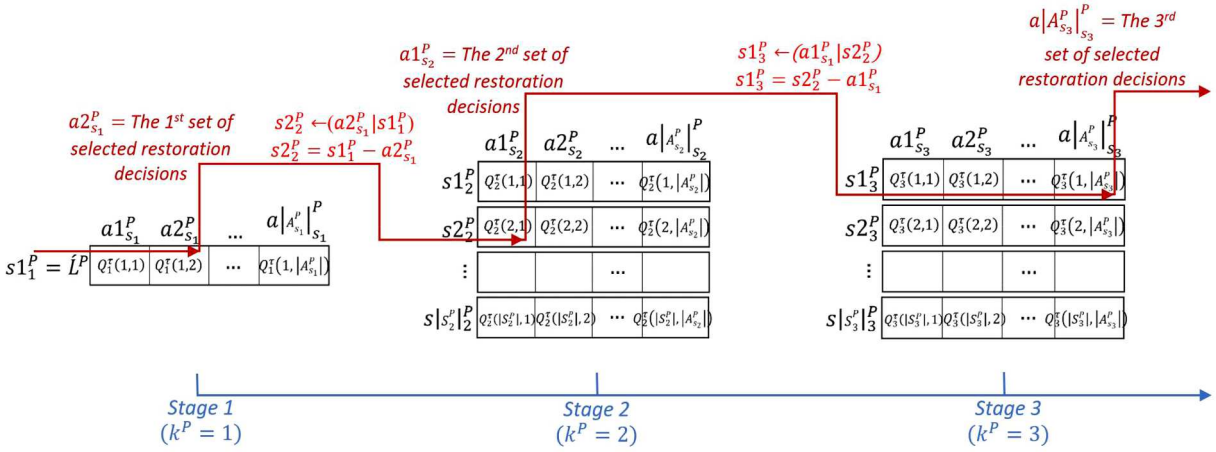


Fig. 3. The sequence of decisions made by the agent in the power MDP.

"coordinator" results in the coupled RL-OPs mechanism that is able to make coordinated restoration decisions for the power and road CIs.

4.1. RL-OP development for the power CI

In this section, we explain the procedure for developing a distinct RL-OP for the power CI. The decision-making process for the agent of the power RL-OP and the reward generation process in its learning environment are explained in Sections 4.1.1 and 4.1.2, respectively. Section 4.1.3 details the solution approach for the power MDP, including the learning procedure used by the agent of the power RL-OP to train and ultimately generate uncoordinated restoration policies for the disrupted power network.

4.1.1. Decision-making process for the agent of the power RL-OP

As explained before, in the presence of limited recovery resources (e.g., limited recovery teams), the agent of the power RL-OP should determine the best sequence and concurrency for recovering disrupted links in the power network. Therefore, we formulate the problem of restoration scheduling and resource allocation by the agent as a MDP with sequential decision-making stages.

In the context of the power network, let $\acute{L}^P$ be the set of disrupted links. At each decision-making stage of the power MDP ($\forall k^P \in K^P$), the agent is tasked with two key decisions: (1) selecting a subset of disrupted links for restoration ($L^{'P}_k \subset L^{'P}$); and (2) allocating recovery teams (Λ^P_k) available at stage k^P to the links selected for restoration in that stage. The number of recovery teams (Λ^P_k) is dynamic and may change from stage to stage. The process of allocating recovery teams to a subset of links (e.g., $L^{'P}_k$) that can be selected for restoration at stage k^P is explained the next section, addressing resource allocation decisions. The process of modeling $L^{'P}_k$ selection in each stage of the MDP will be explained in the later section, addressing restoration scheduling decisions (the notation used in the paper is summarized in Table A1 in Appendix A.).

4.1.1.1. Resource allocation decisions. The recovery time needed to restore each link in set $L^{'P}_k$ depends on the number of recovery teams allocated to that link. The resource allocation process aims to minimize the restoration time for the selected links in $L^{'P}_k$, ensuring their swift recovery. We develop an optimization model to assign recovery teams to the links of set $L^{'P}_k$. In this model, variable w^P_l indicates the number of teams allocated to link $l \in L^{'P}_k$, while $\overline{\sigma}_l$ demonstrates the average restoration time for link l if only one team were allocated to it (where σ_l is a random variable ranging from $\check{\sigma}_l$ to $\hat{\sigma}_l$). The optimal allocation of teams to the links in set $L^{'P}_k$ is determined by Model (1–4):

$$\text{Minimize } \vartheta_{L^{'P}_k} = \underset{\forall l \in L^{'P}_k}{\text{MAX}} \vartheta^l = \left(\frac{\overline{\sigma}_l}{w^P_l}\right) \tag{1}$$

$$\text{Subject to : } w^P_l \le C_l \ \forall l \in L^{'P}_k \tag{2}$$

$$\sum_{l \in L^{'P}_k} w^P_l \le \Lambda^P_k \tag{3}$$

$$w^P_l \ge 0 \text{ and integer } \forall l \in L^{'P}_k \tag{4}$$

Objective function (1) allocates recovery teams to the links of set $L^{'P}_k$ in a way to minimize the maximum time needed to recovery of each link in that set. It is assumed that the allocated teams will remain dedicated to their assigned links throughout the recovery process, and the restoration of links in $L^{'P}_k$ will be considered complete once all links are restored. This assumption is commonly adopted in literature (Çelik et al., 2015; Averbakh, 2012; Tzeng et al., 2007). Constraint (2) ensures that the number of allocated teams to each link does not exceed the maximum number of teams that can work simultaneously on that link (C_l). Additionally, constraint (3) guarantees that the number of allocated teams at stage k does not exceed the total number of available teams at that stage (Λ^P_k). The process of linearizing and solving Model (1–4) is explained in Appendix B.

4.1.1.2. Restoration scheduling decisions. This section explains the process of modeling $L^{'P}_k$ selection by the agent of the power RL-OP at each stage of its MDP. As depicted in Fig. 3, the MDP consists of several decision-making stages. Each stage corresponds to a state-decision matrix. The states (or rows) of a matrix represent potential configurations that the power network may have at the beginning of its corresponding stage. The decisions (or columns) of the matrix represent the potential restoration decisions that can be selected by the agent. The remainder of this section explains the interconnections among the decisions made at different stages of the MDP.

In the state-decision matrix of stage k, the initial configuration of the power network is represented by set $S^P_k = \{s^P_k\}$. Each state ($\forall s^P_k \in S^P_k$) corresponds to a set of disrupted links that remain unrecovered up to that stage. In the first stage ($k^P = 1$), there exists a single state encompassing all links in $L^{'P}$ available for restoration (see Stage 1 in Fig. 3). The decision space in state s^P_k, denoted by $A^P_{s_k} = \{a^P_{s_k}\}$, encompasses all feasible subsets of links that can be selected for restoration in that state. Decision $a^P_{s_k}$ is feasible if solving Model (1–4) yields a finite minimum recovery time $\vartheta^*_{L^{'P}_k = a^P_{s_k}}$ and the optimal team allocation scheme

$\{w_l^{P*}\}_{L_k'^P=a_{s_k}^P}$ for the selected links in that decision. In the initial stage ($k^P = 1$), the decision space includes a maximum of $2^{|L'^P|}$ decisions. However, depending on the availability of recovery teams ($\Lambda_{k=1}^P$), some of these decisions may become infeasible. In the case in which decision $a_{s_1=}^P L_1'^P$ is selected by the agent in stage 1 for restoration, the pool of available links for restoration in stage 2 diminishes to $L'^P - L_1'^P$, consequently reducing the maximum size of the decision space to $2^{|L'^P - L_1'^P|}$.

As shown by red arrows in Fig. 3, the initial state of the power CI in stage $k+1$ (s_{k+1}^P) depends on its initial state in stage k (s_k^P) and the restoration decision made by the agent in stage k $\left(a_{s_k}^P\right)$: $s_{k+1}^P \leftarrow (a_{s_k}^P | s_k^P)$. This transition function interconnects decisions made in sequential stages of an MDP. The reward of selecting decision $a_{s_k}^P$ in state s_k, represented by $\theta_k^P\left(a_{s_k}^P\right)$, is calculated based on the total increase that making this decision (recovering the selected links) will make in fulfilling power demand of the community from the moment that the restoration for decision $a_{s_k}^P$ ends up to T. This reward will be calculated in the learning environment of the power RL-OP that will be elaborated in Section 4.1.2.

MDP is one of the most well-known approaches for making sequential decisions in stochastic environments. That is why it is used to frame the decision-making process for the agents of RL-OPs. It can handle uncertainties that may arise in the implementation process of selected decisions, which affect transition functions ($s_{k+1}^P \leftarrow (a_{s_k}^P | s_k^P)$) (Nilim and Ghaoui, 2005) and uncertainties that may occur in rewards generated after implementing decisions ($\theta_k^P\left(a_{s_k}^P\right)$) (Paschalidis & Kang, 2008). In this paper, we focus only on uncertainties in recovery times that impact rewards. However, employing MDPs provides the opportunity to consider other types of uncertainties in the problem formulation.

4.1.2. Learning environment of the power RL-OP

Assume that decision $a_{s_k}^P$ is selected at stage k (corresponds to the decision-making moment of t^k), and the total time needed to restore the links of this decision is $\vartheta^*_{L_k'^P=a_{s_k}^P}$, calculated by Model (1–4). The reward of making this decision, $\theta_k^P\left(a_{s_k}^P\right)$, would be equal to the total power demand that can be fulfilled by the power CI in $[t^k + \vartheta^*_{a_{s_k}^P},\ T]$ interval in the presence of links of set $a_{s_k}^P$ minus the total power demand that can be fulfilled by the power CI in $[t^k + \vartheta^*_{a_{s_k}^P},\ T]$ interval in the absence of links of set $a_{s_k}^P$. The demand that can be fulfilled by the power CI with a given set of active links in each time unit (e.g., a day) is calculated using Model (5–13). This model optimizes the power distribution in the power CI, $G^P(N^P, L^P)$, under different link availability scenarios. To include link availability scenarios, we will assign a binary parameter (β_l^P) to each link in the power network. Parameter $\beta_l^P = 1$ if directed link $l = (\overrightarrow{n,n}) \in L^P$ is active and can be employed for transferring power in the power network, and 0 otherwise.

$$\text{Minimize } Z^P = \sum_{n \in N_D^P} UD_n^P \tag{5}$$

$$\text{Subject to : } \sum_{n' \in N^P} x^P_{l=(\overrightarrow{n,n'})} \le PC_n^P \ \left(\forall n \in N_S^P\right) \tag{6}$$

$$\sum_{n' \in N^P} x^P_{l=(\overrightarrow{n',n})} = \sum_{n'' \in N^P} x^P_{l=(\overrightarrow{n,n''})} \ \left(\forall n \in N_I^P\right) \tag{7}$$

$$\sum_{n' \in N^P} x^P_{l=(\overrightarrow{n',n})} = DD_n^P - UD_n^P \ \left(\forall n \in N_D^P\right) \tag{8}$$

$$x^P_{l=(\overrightarrow{n',n})} \le TC^P_{l=(\overrightarrow{n',n})} . \beta^P_{l=(\overrightarrow{n',n})} . y^P_{l=(\overrightarrow{n',n})} \ \left(\forall l \in L^P\right) \tag{9}$$

$$y^P_{l=(\overrightarrow{n',n})} + y^P_{l=(\overrightarrow{n,n'})} \le 1 \ \left(\forall l \in L^P\right) \tag{10}$$

$$b_l^P . x^P_{l=(\overrightarrow{n',n})} \le (\varphi_{n'} - \varphi_n) + M\left(1 - y^P_{l=(\overrightarrow{n',n})}\right) \ \left(\forall l \in L^P\right) \tag{11}$$

$$b_l^P . x^P_{l=(\overrightarrow{n',n})} \ge (\varphi_{n'} - \varphi_n) - M\left(1 - y^P_{l=(\overrightarrow{n',n})}\right) \ \left(\forall l \in L^P\right) \tag{12}$$

$$x_l^P, b_l^P, \varphi_n \ge 0 \text{ and } y_l^P \in \{0,1\} \ \left(\forall n \in N^P\right) \ \left(\forall l \in L^P\right) \tag{13}$$

Objective function (5) minimizes the total unfulfilled demand at the demand nodes of the power CI during a day. Variable UD_n^P measures the daily demand that cannot be fulfilled at node $n \in N_D^P$ under the link availability scenario of $\{\beta_l^P | \forall l \in L^P\}$. Therefore, the maximum demand that can be fulfilled per day is equal to $\sum_{n \in N_D^P} DD_n^P - Z^{*P}$. Based on constraint (6), the total power flow originating from a supply node cannot violate its generation capacity (PC_n^P). At intermediary nodes, the sum of power inflow must equal the sum of power outflow (constraint (7)). At each demand node, the total power inflow is equal to the fulfilled portion of the demand at that node (constraint (8)). Constraint (9) ensures that the power flows only through the links available in that scenario and in the movement direction identified by variable $y^P_{l=(\overrightarrow{n',n})}$. Variable $y^P_{l=(\overrightarrow{n',n})}$ is 1 if the movement direction of power is from node n' toward node n, and 0 otherwise. Through each link, the power flow is only possible in one direction (constraint (10)). Constraints (11) and (12) are related to physics of the power network that is approximated as a linear DC model. In the DC model, the power flow through each link should be consistent with the reactance of that link (b_l^P) and the phase angle of its connecting nodes (φ_n and $\varphi_{n'}$). For more details, refer to Nurre et al. (2012).

Model (5–13) optimizes the power distribution in the power network under the link availability scenario of $\{\beta_l^P | \forall l \in L^P\}$. This optimization model constitutes the learning environment of the power RL-OP and provides reward θ_k^P for the restoration decisions $a_{s_k}^P$ made by the agent of the power RL-OP. The process of calculating rewards based on the outcomes of Model (5–13) is detailed in Section 4.1.3.

4.1.3. Solution approach: the learning and optimal policy generation by the agent in the power RL-OP

To link the rewards of decisions made across different power MDP stages, counter-cumulative improvements, denoted as Q values, are calculated for the cells of its state-decision matrices. The Q values in the power MDP of Fig. 3 represent the best counter cumulative rewards achievable by making each decision in each state of the state-decision matrices. For example, $Q\left(s_k^P, a_{s_k}^P\right)$ quantifies the maximum expected improvement attainable from stage k to the final stage $|K|$ if decision $\overrightarrow{a^k}$ is selected by the agent in state s_k^P at stage k. To reduce the computational complexity of calculating Q values of all matrices, RL is used in this paper to estimate Q values. We explain the process of estimating Q values and training the power agent in this section.

In each iteration of power RL-OP, the first restoration decision is made by the agent at time 0 ($k^P = 1$), and a subset of disrupted links, $a_{sk=1}^P = L_{k=1}'^P$ is selected for recovery using the ε-greedy approach (Jasmin et al., 2011). The restoration process of these links will end at time $\vartheta^*_{a_{s_{k=1}}^P = L_{k=1}'^P}$, if average recovery times are materialized for the links of set $L_{k=1}'^P$. To consider uncertainty in the recovery times, random values from the variation ranges of $\left[\check{\sigma}_l - \hat{\sigma}_l\right]$ will be assigned to each link, denoted as $\tilde{\sigma}_l$. These random values are used to calculate the actual time it may take to restore the selected links ($\vartheta' a_{s_{k=1}}^P = L_{k=1}'^P$):

$$\vartheta'_{L'^{P}_{k=1}} = \underset{\forall l \in L'^{P}_{k=1}}{\mathrm{MAX}} \left\{ \frac{\widetilde{\sigma}_l}{w^{P*}_l} \right\} \tag{14}$$

where w^{P*}_l values are calculated by Model (1–4). Selecting disrupted links of $a^P_{s_{k=1}} = L'^P_{k=1}$ for recovery will make some improvement in the daily demand that can be fulfilled by the power CI. This improvement, $\theta^P_{k=1}\left(a^P_{s_{k=1}} = L'^P_{k=1}\right)$, is calculated using Model (5–13) in the RL-OP's learning environment as follows:

$$\theta^P_{k=1}\left(a^P_{s_{k=1}}\right) = Z^{P*}\left(\beta^P_l = 0 \ (\forall l \in L'^P),\ \beta^P_l = 1 \ (\forall l \in L^P - L'^P)\right).\left(T - \vartheta'_{L'^P_{k=1}}\right) - Z^{P*}\left(\beta^P_l = 0 \ (\forall l \in L'^P - L'^P_{k=1}),\ \beta^P_l = 1 \ (\forall l \in L^P - L'^P + L'^P_{k=1})\right).\left(T - \vartheta'_{L'^P_{k=1}}\right) \tag{15}$$

This reward will be used to update the Q estimation of cell $\left(s^P_{k=1}, a^P_{s_{k=1}}\right)$ in the power MDP using the Bellman's equation (Sutton & Barto, 1999):

$$Q^{\tau+1}\left(s^P_{k=1}, a^P_{s_{k=1}}\right) = (1-\alpha) Q^{\tau}\left(s^P_{k=1}, a^P_{s_{k=1}}\right) + \alpha\left[\theta^P_{k=1}\left(a^P_{s_{k=1}}\right) + \gamma \underset{a^P_{s_{k=2}}}{max} Q^{\tau}\left(s^P_{k+1=2}, a^P_{s_{k+1=2}}\right)\right] \tag{16}$$

where α and γ respectively control the convergence speed of the learning process and the weight of future rewards. The selected decision is stage 1 determines the state of the network in stage 2: $s^P_{k=2} \leftarrow (a^P_{s_{k=1}} | s^P_{k=1})$. Then, the second restoration decision is selected by the agent from the action set of state $s^P_{k=2}$ in the second stage ($k^P = 2$) which includes another subset of disrupted links, $L'^P_{k=2}$, that have not been recovered by time $\vartheta'_{L'^P_{k=1}}$. The reward of making this decision and the Q estimation of the selected cell, $\left(s^P_{k=2},\ a^P_{s_{k=2}} = L'^P_{k=2}\right)$, are calculated using the same approach. The agent continues this decision-making process until all disrupted links are restored. This is the end of the first learning iteration.

Using the ε-greedy approach, in different iterations of the RL-OP, different decisions are selected by the agent in MDP stages, and Q values are updated continuously. After a high number of iterations ($\tau \to \infty$), the Q values of the power MDP, estimated by Eq. (16), converge to their actual values. After convergence, the agent derives the optimal link restoration policy for the power CI ($\pi^{P*} : S^P \to A^P$) as follows:

$$\pi^{P*} = Argmax_{\pi^{P*}}\left(\sum_{k^P=1}^{|K^P|} \gamma^{k^P}.\theta^P_k\right) \tag{17}$$

This policy recommends the best restoration schedule and resource allocation for the disrupted links of the power CI with stochastic recovery times and dynamic number of recovery teams. However, the interdependencies of these decisions to the restoration decisions of the road CI are completely ignored in this policy. Therefore, some restoration decisions of the recommended policy may not be executable in practice.

4.2. RL-OP development for the road CI

This section explains the process of developing a distinct RL-OP for the road CI. Section 4.2.1 explains the decision-making process for the agent of the road RL-OP. The consequence/reward of the decisions made by the agent is evaluated in the road RL-OP's learning environment, which will be elaborated in Section 4.2.2. The learning procedure for the agent which helps it generate an uncoordinated restoration policy for the disrupted links of the road network is explained in Section 4.2.3.

4.2.1. Decision-making process for the agent of the road RL-OP

Similar to the power RL-OP, the problem of identifying the best restoration schedule and team allocation for the disrupted links in the road network is formulated as an MDP. Assuming that L'^R is the set of disrupted links in the road network, at each decision-making stage of the road MDP ($\forall k^R \in K^R$), the agent selects a subset of disrupted links for restoration ($L'^R_k \subset L'^R$) and allocates recovery teams (Λ^R_k) to those links ($w^R_l, l \in L'^R_k$). The method used for team allocation is similar to the power RL-OP. Model (1–4) is used to determine the best pattern of assigning recovery teams to the links selected for simultaneous restoration in L'^R_k, $\left\{w^{R*}_l\right\}_{L'^R_k}$.

The decision-making procedure used by the agent of the road RL-OP for the restoration schedule is the same as for the power RL-OP. At each stage of the road MDP, the initial configuration of the road network is represented by a set of potential states: $S^R_k = \left\{s^R_k\right\}$. Each state corresponds to a set of disrupted links that have not been recovered up to that stage. In state s^R_k, the decision space, denoted as $A^R_{s_k} = \left\{a^R_{s_k}\right\}$, encompasses all feasible subsets of links eligible for restoration. The reward associated with selecting decision $a^R_{s_k}$ in state s_k is determined by quantifying the overall reduction in post-disaster travel time/cost resulting from making decision $a^R_{s_k}$ (i.e., restoring the chosen roads of this decision). This reduction will be calculated from the moment that the restoration operation ends for decision $a^R_{s_k}$ up to T. This reward is calculated in the learning environment of the road RL-OP, which will be elaborated in Section 4.2.2.

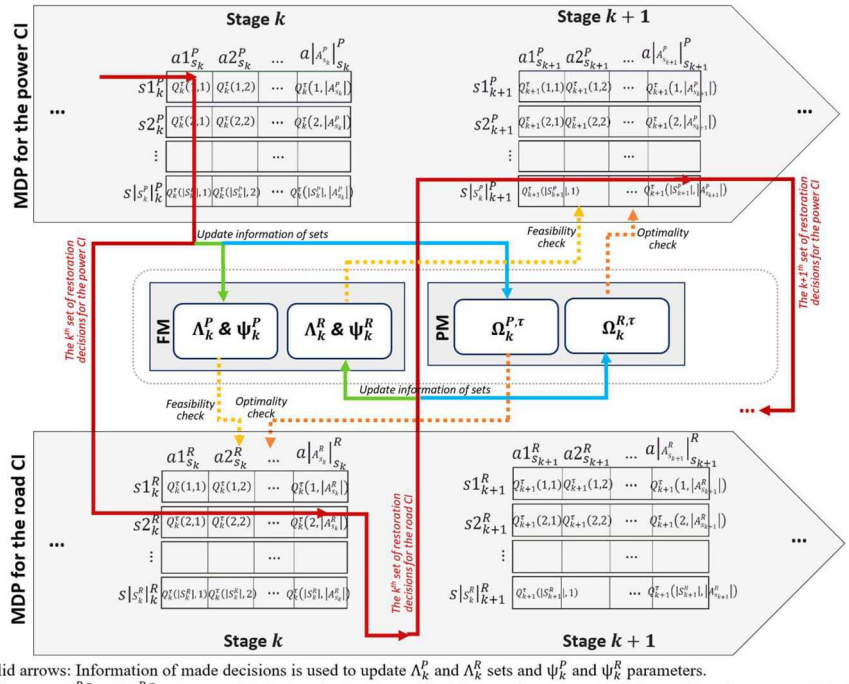
4.2.2. Learning environment of the road RL-OP

In the case in which decision $a^R_{s_k}$ is selected at stage k (corresponds to the decision-making moment t^k), the time needed to complete this restoration operation is $\vartheta'_{a^R_{s_k}}$. This means the links of decision $a^R_{s_k}$ will be available for use at time $t^k + \vartheta'_{a^R_{s_k}}$. So, the reward of making this decision, $\theta^R_k\left(a^R_{s_k}\right)$, would be equal to the total travel time/cost in the road CI in $[t^k + \vartheta'_{a^R_{s_k}},\ T]$ interval in the absence of the links of set $a^P_{s_k}$ minus the total travel time/cost in the road CI in $[t^k + \vartheta'_{a^R_{s_k}},\ T]$ interval in the presence of links of set $a^P_{s_k}$. The post-disaster travel time/cost in the road CI in each time unit (e.g., each day) is calculated using Model (18–23). This model optimizes the traffic routing over the road CI, $G^R\left(N^R, L^R\right)$, under different link availability scenarios. The availability of links in the network is determined by a binary parameter β^R_l. Parameter $\beta^R_l = 1$ if link $l = (n, n')$ is active and available for traveler usage, and $\beta^R_l = 0$ otherwise.

$$\text{Minimize } Z^R = \sum_{\forall od \in OD} \sum_{\forall l \in L^R} x^{R,od}_{l=(n,n')}.tt_{l=(n,n')} \tag{18}$$

$$\text{Subject to : } \sum_{\forall od \in OD} x^{R,od}_{l=(n',n)} \le FC^R_{l=(n',n)}.\beta^R_{l=(n',n)} \left(\forall l \in L^R\right) \tag{19}$$

$$\sum_{n \in N^R} x^{R,od}_{l=(m,n)} = TF^R_{od} (\forall od = (m, m') \in OD) \tag{20}$$



Green solid arrows: Information of made decisions is used to update Λ_k^P and Λ_k^R sets and ψ_k^P and ψ_k^R parameters.
Blue solid arrows: $\Omega_k^{P,\tau}$ and $\Omega_k^{R,\tau}$ sets are updated when the Q values of decision matrixes are revised after making the new set of decisions.
Yellow dashed arrows: Before making a new set of decisions, the feasibility of decisions is checked according to Λ_k^P and Λ_k^R sets and ψ_k^P and ψ_k^R parameters.
Orange dashed arrows: Before making a new set of decisions, the optimality of decisions is checked according to $\Omega_k^{P,\tau}$ and $\Omega_k^{R,\tau}$ sets.
Red solid arrows: Show the sequence of decisions among the decision matrixes of power and road MDPs.

Fig. 4. The decision-making structure of the coupled RL-OPs.

$$\sum_{n\in N^R} x_{l=(n,m')}^{R,od} = TF_{od}^R (\forall od=(m,m')\in OD) \tag{21}$$

$$\sum_{n\in N^R} x_{l=(n,n')|n'\neq m,m'}^{R,od} = \sum_{n''\in N^R} x_{l=(n',n'')|n'\neq m,m'}^{R,od} (\forall od=(m,m')\in OD) \tag{22}$$

$$x_l^{R,od} \geq 0 (\forall l\in L^R) \text{ and } (\forall od\in OD) \tag{23}$$

Objective function (18) minimizes the total post-disaster travel time/cost in the road network. Parameter tt_l represents the travel time/cost for a traveler moving through link $l\in L^R$. According to constraint (19), traffic can only flow through active links (when $\beta_l^R = 1$), and the flow volume cannot exceed the capacity of the link (FC_l^R). Constraint (20) ensures that the total traffic outflow from the origin node of each OD pair is equal to the traffic demand of that OD. Similarly, the total traffic inflow to the destination node of each OD is equal to the OD's traffic demand (constraint (21)). At intermediate nodes, which are neither the origin nor the destination of an OD pair, the sum of inflow must equal the sum of outflow, as expressed by constraint (22).

As explained before in Section 3.2, we assume that routine pre-disaster traffic flows that may cause traffic jams in roads do not exist in the area. In the cases in which we expect traffic jams (more than travel flow capacity) and delayed travel time in the links of the road CI, we can replace parameter tt_l (that shows the travel time/cost for a traveler moving through link $l\in L^R$) with a function that connects the travel time/cost of a link to its capacity and traffic flow. One of these functions that is suggested by the Bureau of Public Roads (Bureau of Public Roads, Anon., 1964) and widely used in the literature is:

$$tt_l\left(\sum_{\forall od\in OD} x_l^{R,od}\right) = c_0\left[1+0.15\left(\frac{\sum_{\forall od\in OD} x_l^{R,od}}{FC_l^R}\right)^4\right]$$

In this function, traversing link l is associated with a positive cost/time of tt_l for travelers, which is a function of its traffic flow ($\sum_{\forall od\in OD} x_l^{R,od}$), free-flow travel time (c_0), and nominal capacity (FC_l^R). Also, we need to modify constrain (19) in Model (18–23) as: $\sum_{\forall od\in OD} x_{l=(n',n)}^{R,od} \leq M.\beta_{l=(n',n)}^R$ where M is a very large positive value.

Model (18–23) determines the best post-disaster traffic pattern in the road CI with the minimum travel cost/time under the link availability scenario of $\{\beta_l^R | \forall l\in L^R\}$. As the learning environment of the road RL-OP, this model provides rewards for the restoration decisions made by the agent of the road RL-OP at each stage of its MDP. The process of calculating rewards based on the outcomes of Model (*18–23*) is detailed in Section 4.2.3.

4.2.3. Solution approach: the learning and optimal policy generation by the agent in the road rl-op

The RL mechanism, similar to the power CI, is used in this section to estimate Q values for state-decision matrices of the road MDP. In each learning iteration, the first set of restoration decisions ($k^R = 1$) is made at time 0, and a subset of disrupted links, $a_{s_{k=1}}^R = L_{k=1}^{'R}$, is selected by the agent for recovery using the ε-greedy approach. The restoration process of these links will end at time $\vartheta'_{L_{k=1}^{'R}}$. This decision will make some improvement in daily traffic time/cost throughout the road CI, which will be calculated using Model (18–23) in the RL-OP's learning environment. The outcomes of the model used to calculate the decision's reward as follows:

$$\theta_{k=1}^R\left(a_{s_{k=1}}^R\right) = Z^{R*}\left(\beta_l^R = 0\left(\forall l\in L^{'R}\right), \beta_l^R = 1\left(\forall l\in L^R - L^{'R}\right)\right).\left(T-\vartheta'_{L_{k=1}^{'R}}\right) - Z^{R*}\left(\beta_l^R = 0\left(\forall l\in L^{'R} - L_{k=1}^{'R}\right), \beta_l^R = 1\left(\forall l\in L^R - L^{'R} + L_{k=1}^{'R}\right)\right).\left(T-\vartheta'_{L_{k=1}^{'R}}\right) \tag{24}$$

Eq. (24) calculates the total reduction in travel time/cost during T if we select to restore links of set $a_{s_{k=1}}^R$ at $k^R = 1$. This reward is used to update the Q estimation of cell $\left(s_{k=1}^R, a_{s_{k=1}}^R\right)$ in the road MDP as follows:

$$Q^{\tau+1}\left(s^R_{k=1}, a^R_{s_{k=1}}\right) = (1-\alpha)Q^{\tau}\left(s^R_{k=1}, a^R_{s_{k=1}}\right) + \alpha\left[\theta^R_{k=1}\left(a^R_{s_{k=1}}\right) + \gamma \underset{a^R_{s_{k=2}}}{maxQ^{\tau}}\left(s^R_{k=2}, a^R_{s_{k=2}}\right)\right] \quad (25)$$

The similar procedure is repeated to update Q values for other stages of the MDP. After a high number of iterations ($\tau \to \infty$), the Q values of the road MDP converge to their actual values. After convergence, the agent derives the optimal link restoration policy for the road CI ($\pi^{R*}: S^R \to A^R$) as follows:

$$\pi^{R*} = \underset{\pi^{R*}}{Argmax}\left(\sum_{k^R=1}^{|K^R|} \gamma^{k^R}.\theta^R_k\right) \quad (26)$$

This policy determines the best uncoordinated restoration schedule and resource allocation for the disrupted links of the road CI.

4.3. Coupled RL-OPs for coordinated restoration of road and power CIs

4.3.1. Coordinator development

There is no communication or information sharing between the agents of the power and road RL-OPs developed in Sections 4.1 and 4.2. This lack of coordination results in overlooking procedural interdependencies (represented by sets Π^P_l ($\forall l \in L^P$) and Π^R_l ($\forall l \in L^R$)) in the restoration policies generated by their RL-OPs. Therefore, the uncoordinated policies generated by distinct RL-OPs may be infeasible or suboptimal. Infeasibility means the recommended policy for a CI cannot be executed in practice as is because the prerequisites of the links selected for restoration in each decision-making stage may not have been restored in the interdependent CI. To facilitate communication between the agents of RL-OPs, we will design a "coordinator" that enables partial information sharing between the agents in the coupled RL-OPs. This coordinator prevents them from making infeasible restoration decisions and coordinates their decisions to be more beneficial for the entire community, not just their own CI. The coordinator includes two modules (see Fig. 4):

- **Feasibility Module (FM):** After accomplishing each restoration decision in each RL-OP, the information of the recovered links and their availability times are stored in the FM module (represented by sets Λ^P_k and Λ^R_k and parameters ψ^P_k and ψ^R_k). This information will be available for the agents of other RL-OPs and will prevent them from making infeasible restoration decisions.
- **Prediction Module (PM):** This module shares the maximum Q value that is achievable in each state of each MDP stage with agents of other RL-OPs (represented by arrays $\Omega^{P,\tau}_k$ and $\Omega^{R,\tau}_k$). This information helps those agents predict the consequences of their decisions in each stage on interdependent CIs and select decisions that result in better aggregated rewards (summation of rewards achieved by the agent and the agents of its interdependent CIs) rather than individual rewards.

This means the agent of each RL-OP does not have any information about the network structure (e.g., the number and location of nodes and links in networks), operational limitations (e.g., supply capacities and demand quantities), and restoration operations (e.g., number of recovery teams and recovery times of disrupted links) of other networks. They only have access to the limited information shared through the coordinator. Fig. 4 demonstrates how the decision-making structure of the agents should be modified in the proposed coupled RL-OPs mechanism. There is no change in the RL-OPs' learning environments. In the rest of this section, we explain the flow of information among the state-decision matrices of RL-OPs and the modules of the coordinator:

- *Information sharing through FM*: After making a restoration decision in each stage of an RL-OP, the information of the recovered links and their recovery accomplishment times are recorded in the FM module of the "coordinator" (depicted as solid green arrows in Fig. 4). In the under-study problem of coordinated restoration planning for power and road CIs, there are two RL-OPs. Therefore, the FM module will include two sets:

$$\Lambda^P_k = \left\{\left(l^P, ct^P_l\right)\right\} \text{ and } \Lambda^R_k = \left\{\left(l^R, ct^R_l\right)\right\} \quad (27)$$

Set Λ^P_k includes all the links that have been restored in the power CI up to the stage k^P (l^P) and their recovery accomplishment times (ct^P_l). The same information is recorded in set Λ^R_k for the road CI. Also, we need to keep a record of the decision-making moments throughout the stages of the MDPs:

$$\psi^P_k = \psi^P_{k-1} + \vartheta^{\cdot}_{L^{\cdot P}_{k-1}}\left(1 < k^P \le |K^P|\right) \text{ and } \psi^R_k = \psi^R_{k-1} + \vartheta^{\cdot}_{L^{\cdot P}_{k-1}}\left(1 < k^R \le |K^R|\right) \quad (28)$$

Parameter ψ^P_k represents the time at which the restoration decisions are made at stage k^P of the power MDP. ψ^R_k demonstrates the same time for the road MDP ($\psi^P_1 = \psi^R_1 = 0$).

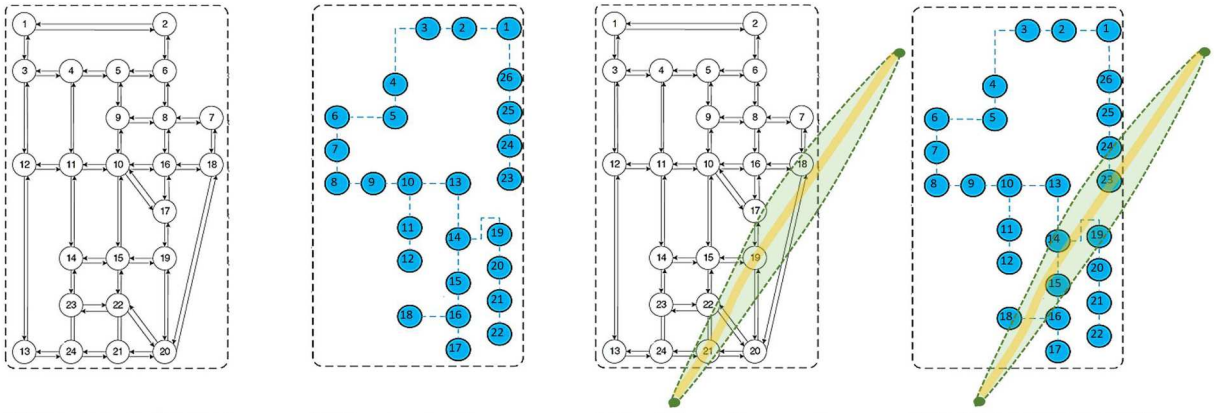
- *Feasibility checking through FM*: Before making any decision in each stage of an MDP, the feasibility of these decisions will be checked with the sets of the FM (depicted as dashed yellow arrows in Fig. 4). For example, the decision $a^R_{s_k}$ is feasible in stage k^R of the road MDP if all of its prerequisites in set $\underset{l \in a^R_{s_k}}{\cup} \Pi^R_l$ exist in set Λ^P_k, and their recovery accomplishment times are less than or equal to ψ^R_k:

$$ct^P_l \le \psi^R_k \text{ for } \forall l \in \underset{l \in a^R_{s_k}}{\cup} \Pi^R_l \quad (29)$$

The sets of feasible decisions in each stage of the power and road MDPs are represented by $\widehat{A}^P_{s_k} = \left\{\widehat{a}^P_{s_k}\right\}$ and $\widehat{A}^R_{s_k} = \left\{\widehat{a}^R_{s_k}\right\}$.

- *Consequence predicting through PM*: The PM of the coordinator helps the agent of each RL-OP to predict the consequences of its decisions on its interdependent CIs. This guides the agent to make coordinated, rather than distinct, decisions because it will consider the impacts of decisions not only on its own CI but also on the interdependent CIs. For this purpose, the PM records the maximum Q value that is achievable in each state of each MDP stage (depicted as solid blue arrows in Fig. 4). These values are not fixed and updated in the iterations of RLs (τ):

$$\Omega^{P,\tau}_k = \left[MQ^{P,\tau}\left(s^P_k\right) = \underset{a^P_{s_k}}{\max} Q^{\tau}\left(s^P_k, a^P_{s_k}\right)\right] \text{ and } \Omega^{R,\tau}_k = \left[MQ^{R,\tau}\left(s^R_k\right) = \underset{a^R_{s_k}}{\max} Q^{\tau}\left(s^R_k, a^R_{s_k}\right)\right] \quad (30)$$



(a) The network of the road CI. (b) The network of the power CI. (c) Roads located in a sample tornado movement trajectory. (d) Cables located in a sample tornado movement trajectory.

Fig. 5. The road and power CIs of the study region.

The information of these sets is shared with the agents of all RL-OPs (partial information sharing) to help them select coordinated decisions (depicted as dashed orange arrows in Fig. 4). To employ these predictions in the decision-making process of agents, we revise the Bellman's equation, Eq. (16), as follows:

$$Q^{\tau+1}\left(s_k^P, a_{s_k}^P\right) = (1-\alpha)Q^{\tau}\left(s_k^P, a_{s_k}^P\right) + \alpha\left[\theta_k^P\left(a_{s_k}^P\right) + \gamma \underset{a_{s_{k+1}}^P}{max} Q^{\tau}\left(s_{k+1}^P, a_{s_{k+1}}^P\right) + \lambda MQ^{R,\tau}\left(s_k^R\right)\right] \quad (31)$$

According to (31), the agent of the power RL-OP not only considers the impact of the decisions made at stage k^P on the future achievable rewards in the power CI, term $\underset{a_{s_{k+1}}^P}{max} Q^{\tau}\left(s_{k+1}^P, a_{s_{k+1}}^P\right)$, but also considers the sequence of these decisions on the maximum achievable reward in the road CI, term $MQ^{R,\tau}\left(s_k^R\right)$. Parameter λ represents the importance of the road CI performance for the agent of the power RL-OP. Similarly, the Q values for the road CI's agent will be calculated as follows:

$$Q^{\tau+1}\left(s_k^R, a_{s_k}^R\right) = (1-\alpha)Q^{\tau}\left(s_k^R, a_{s_k}^R\right) + \alpha\left[\theta_k^R\left(a_{s_k}^R\right) + \gamma \underset{a_{s_{k+1}}^R}{max} Q^{\tau}\left(s_{k+1}^R, a_{s_{k+1}}^R\right) + \lambda MQ^{P,\tau}\left(s_k^P\right)\right] \quad (32)$$

For the decision-maker (or agent) of a CI, the priority of rewards achievable within its own CI may be higher than those from interdependent CIs. This is why we included the parameter "λ" in Eqs. (31) and (32). In Eq. (31), this parameter represents the importance of the reward achievable in the road CI for the decision-maker of the power CI. When λ=1, the rewards of both CIs have equal priority for the decision-maker of the power CI. When λ<1, the reward in the power CI has higher priority than that of the road CI. When λ=0, the decision-maker of the power CI prioritizes the individual reward of its own CI.

4.3.2. Reward normalization

When the rewards of CIs are on completely different scales, it is necessary to normalize the rewards generated by the learning environments of the RL-OPs. This ensures that they (and consequently *Q* values) are on a common scale, making them comparable with each other. For example, in the problem of coordinated restoration planning for the power and road CIs, the reward for restoring a given set of links in a stage of the power MDP is calculated as follows:

$$Reward = \left(\begin{array}{c}\text{Total unfulfilled power demand during the remaining}\\ \text{portion of } T \text{ in the absence of the selected links}\end{array}\right) - \left(\begin{array}{c}\text{Total unfulfilled power demand during the remaining}\\ \text{portion of } T \text{ in the presence of the selected links}\end{array}\right)$$

For normalization, we need to adjust the reward calculation in the power MDP as follows:

$$Normalized\ reward = \frac{\left(\begin{array}{c}\text{Total unfulfilled power demand during the remaining}\\ \text{portion of } T \text{ in the absence of the selected links}\end{array}\right) - \left(\begin{array}{c}\text{Total unfulfilled power demand during the remaining}\\ \text{portion of } T \text{ in the presence of the selected links}\end{array}\right)}{\text{Total unfulfilled power demand during } T \text{ in the absence of any restoration activity}}$$

Similarly, the reward for restoring a given set of links in a stage of the road MDP can be normalized as follows:

$$Normalized\ reward = \frac{\left(\begin{array}{c}\text{Total travel cost/time during the remaining}\\ \text{portion of } T \text{ in the absence of the selected links}\end{array}\right) - \left(\begin{array}{c}\text{Total travel cost/time during the remaining}\\ \text{portion of } T \text{ in the presence of the selected links}\end{array}\right)}{\text{Total travel cost/time during } T \text{ in the absence of any restoration activity}}$$

Table 1
Implementation bias of uncoordinated restoration policies generated for interdependent CIs.

Scenario	Features of Scenarios				Policies generated by distinct RL-OPs							Implementation bias		
	Power CI		Road CI		Nominal reward for power CI	Nominal reward for road CI	Nominal reward in both CIs	Feasibility of policies		Actual reward for power CI	Actual reward for Road CI	Actual reward in both CIs	In power CI	In road CI	In both CIs
	# of disrupted links	# of recovery teams	# of disrupted links	# of recovery teams				Power CI	Road CI						
1	3	3	3	3	5800	11,200	17,000	✗	✓	5200	11,200	16,400	11.5 %	0.0 %	3.6 %
2	3	6	3	6	5800	11,600	17,400	✗	✗	5200	11,200	16,400	11.5 %	3.6 %	6.1 %
3	3	9	3	9	5800	11,600	17,400	✗	✗	5200	11,200	16,400	11.5 %	3.6 %	6.1 %
4	7	3	7	3	15,600	23,000	38,600	✗	✗	15,000	21,600	36,600	4.0 %	6.5 %	5.5 %
5	7	6	7	6	15,800	23,220	39,020	✓	✗	15,800	22,030	37,830	0.0 %	5.4 %	3.1 %
6	7	9	7	9	16,800	23,850	40,650	✗	✗	16,500	22,620	39,120	1.8 %	5.4 %	3.9 %
7	9	3	9	3	16,400	42,330	58,730	✗	✗	13,800	33,410	47,210	18.8 %	26.7 %	24.4 %
8	9	6	9	6	18,600	45,400	64,000	✗	✗	18,200	42,800	61,000	2.2 %	6.1 %	4.9 %
9	9	9	9	9	18,800	45,400	64,200	✗	✗	17,700	43,560	61,260	6.2 %	4.2 %	4.8 %
10	10	3	10	3	16,300	62,300	78,600	✗	✗	13,000	56,920	69,920	25.4 %	9.4 %	12.4 %
11	10	6	10	6	18,000	65,530	83,530	✗	✗	16,700	64,500	81,200	7.8 %	1.6 %	2.8 %
12	10	9	10	9	18,700	66,810	85,510	✗	✗	17,300	63,380	80,680	8.1 %	5.4 %	6.0 %
13	11	3	11	3	15,600	127,580	143,180	✗	✗	15,300	118,880	134,180	2.0 %	7.3 %	6.7 %
14	11	6	11	6	17,300	133,930	151,230	✗	✗	16,200	117,390	133,590	6.8 %	14.1 %	13.2 %
15	11	9	11	9	18,500	135,550	154,050	✗	✗	17,600	126,750	144,350	5.1 %	6.9 %	6.7 %
16	12	3	12	3	14,900	141,740	156,640	✗	✗	13,600	105,570	119,170	9.6 %	34.2 %	31.4 %
17	12	6	12	6	18,200	142,140	160,340	✗	✗	17,100	129,850	146,950	6.4 %	9.5 %	9.1 %
18	12	9	12	9	18,300	143,710	162,010	✗	✗	18,000	140,380	158,380	1.7 %	2.4 %	2.3 %

Table 2
Actual rewards of coordinated and uncoordinated restoration policies.

Scenario	Features of Scenarios				Policies generated by distinct RL-OPs			Policies generated by coupled RL-OPs					Improvement percentage			Computational time for coupled RL-OPs
	Power CI		Road CI		Actual reward for power CI	Actual reward for road CI	Actual reward for both CIs	Reward for power CI	Reward for road CI	Aggregated reward for both CIs	Feasibility of policies		Power CI	Road CI	Both CIs	
	# of disrupted links	# of recovery teams	# of disrupted links	# of recovery teams							Power CI	Road CI				
1	3	3	3	3	5200	11,200	16,400	5200	11,200	16,400	✓	✓	+0.0 %	+0.0 %	+0.0 %	5:16:12
2	3	6	3	6	5200	11,200	16,400	5600	11,000	16,600	✓	✓	+7.7 %	−1.8 %	+1.2 %	3:05:04
3	3	9	3	9	5200	11,200	16,400	5600	11,000	16,600	✓	✓	+7.7 %	−1.8 %	+1.2 %	3:13:00
4	7	3	7	3	15,000	21,600	36,600	15,600	21,610	37,210	✓	✓	+4.0 %	+0.0 %	+1.7 %	5:39:36
5	7	6	7	6	15,800	22,030	37,830	16,300	22,220	38,520	✓	✓	+3.2 %	+0.9 %	+1.8 %	5:54:46
6	7	9	7	9	16,500	22,620	39,120	16,600	23,250	39,850	✓	✓	+0.6 %	+2.8 %	+1.9 %	5:52:00
7	9	3	9	3	13,800	33,410	47,210	17,000	40,340	57,340	✓	✓	+23.2 %	+20.7 %	+21.5 %	9:11:17
8	9	6	9	6	18,200	42,800	61,000	18,700	43,420	62,120	✓	✓	+2.7 %	+1.4 %	+1.8 %	8:03:47
9	9	9	9	9	17,700	43,560	61,260	19,100	43,810	62,910	✓	✓	+7.9 %	+0.6 %	+2.7 %	17:39:23
10	10	3	10	3	13,000	56,920	69,920	16,300	60,680	76,980	✓	✓	+25.4 %	+6.6 %	+10.1 %	12:47:42
11	10	6	10	6	16,700	64,500	81,200	18,100	65,570	83,670	✓	✓	+8.4 %	+1.6 %	+3.0 %	11:15:25
12	10	9	10	9	17,300	63,380	80,680	18,600	67,820	86,420	✓	✓	+7.5 %	+7.0 %	+7.1 %	13:22:39
13	11	3	11	3	15,300	118,880	134,180	16,200	122,490	138,690	✓	✓	+5.9 %	+3.0 %	+3.4 %	22:52:46
14	11	6	11	6	16,200	117,390	133,590	17,400	128,540	145,940	✓	✓	+7.4 %	+9.5 %	+9.2 %	23:51:41
15	11	9	11	9	17,600	126,750	144,350	18,100	129,160	147,260	✓	✓	+2.8 %	+1.9 %	+2.0 %	26:06:50
16	12	3	12	3	13,600	105,570	119,170	15,400	137,010	152,410	✓	✓	+13.2 %	+29.8 %	+27.9 %	5 days, 3:47:58
17	12	6	12	6	17,100	129,850	146,950	17,200	140,010	157,210	✓	✓	+0.6 %	+7.8 %	+7.0 %	4 days, 16:05:51
18	12	9	12	9	18,000	140,380	158,380	18,500	148,870	167,370	✓	✓	+2.8 %	+6.0 %	+5.7 %	4 days, 12:05:51

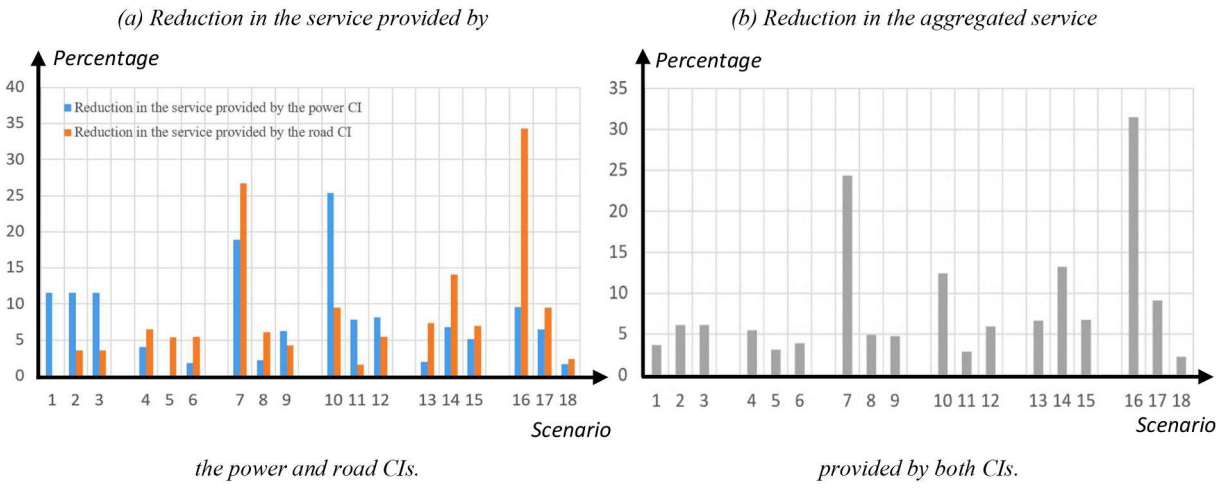


Fig. 6. The implementation bias of the uncoordinated restoration policies.

This normalization process converts rewards into ratios within [0, 1] interval, bringing them to a common scale and making them comparable.

4.3.3. Sequence of decisions in the coupled RL-OPs

In the first learning iteration, the sequence of decisions in the coupled RL-OPs of the road and power CIs is as follows:

- Using the ε-greedy approach and considering $\Lambda^{R}_{k=1} = \varnothing$ in the FM and $\Omega^{R,\tau=1}_{k=1} = \left[\overrightarrow{0}\right]$ in the PM, the agent of the power RL-OP makes the first set of restoration decisions, $\widehat{a}^{P}_{s_{k=1}}$, for the power CI at stage $k^{P} = 1$. After making these decisions, the information of the links selected for recovery and their restoration accomplishment times is added to set $\Lambda^{P}_{k=2}$ in the FM. After updating Q values using Eq. (31), set $\Omega^{P,\tau=1}_{k=1}$ is updated in the PM. Note that the time at the moment of making these decisions is $\psi^{P}_{1} = 0$. After making these decisions, the time increases to $\psi^{P}_{2} = \psi^{P}_{1} + \vartheta'_{\widehat{a}^{P}_{s_{k=1}}}$ in the power RL-OP.
- Then, using the ε-greedy approach and considering $\Lambda^{P}_{k=1} = \varnothing$ in the FM and $\Omega^{P,\tau=1}_{k=1}$ in the PM, the agent of the road RL-OP makes the first set of restoration decisions, $\widehat{a}^{R}_{s_{k=1}}$, for the road CI at stage $k^{R} = 1$. The information of the links selected for recovery and their restoration accomplishment times is added to set $\Lambda^{R}_{k=2}$ in the FM. After updating Q values using Eq. (32), set $\Omega^{R,\tau=1}_{k=1}$ is updated in the PM. Also, the decision-making time increases to $\psi^{R}_{2} = \psi^{R}_{1} + \vartheta'_{\widehat{a}^{R}_{s_{k=1}}}$ in the road RL-OP.
- Then, considering $\Lambda^{R}_{k=2}$ in the FM and $\Omega^{R,\tau=1}_{k=2} = \left[\overrightarrow{0}\right]$ in the PM, the agent of the power RL-OP makes the second set of restoration decisions, $\widehat{a}^{P}_{s_{k=2}}$, for the power CI using the ε-greedy approach. The information of selected links is added to set $\Lambda^{P}_{k=3}$ in the FM, used to calculate new Q values and update set $\Omega^{P,\tau=1}_{k=2}$ in the PM, and employed to increase the power RL-OP time to $\psi^{P}_{3} = \psi^{P}_{2} + \vartheta^{*}_{\widehat{a}^{P}_{s_{k=2}}}$.
- This procedure continues for all stages in the coupled RL-OPs. In the other iterations, all the calculations will be the same, but the Q values of the previous iteration will substitute the zero values of Q in the MDP matrices.

5. Experimental results

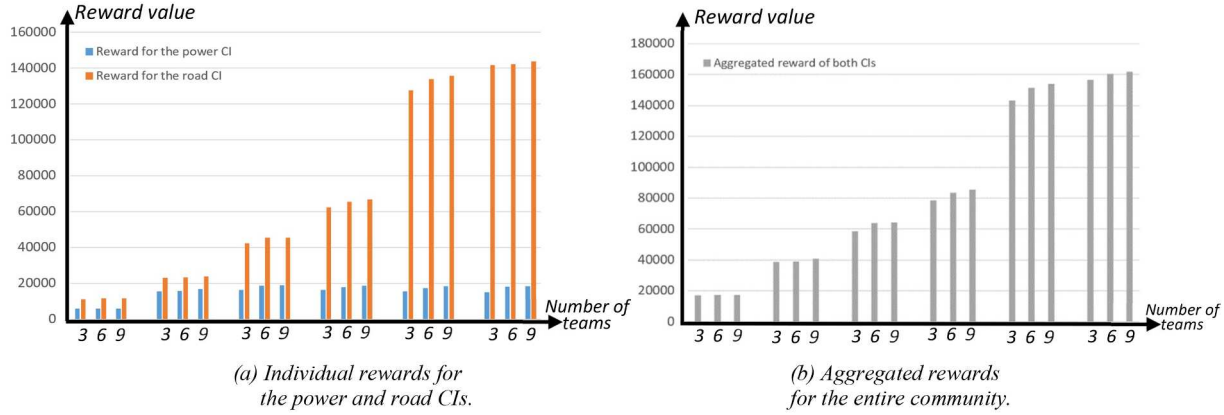
5.1. Study region

Tornadoes are a prevalent natural disaster in the U.S., with an average of 1200 occurrences annually (Perkins, 2002). The U.S. experiences a higher frequency of severe tornadoes, including those categorized as EF4 and EF5, compared to other regions worldwide. Particularly, these severe tornadoes are common in the central U.S., predominantly on the eastern side of the Rocky Mountains. The term "Tornado Alley" is often used to denote the most tornado-prone areas in the U.S., stretching from northern Texas to the Canadian prairies and encompassing several states such as Texas, Louisiana, Oklahoma, Kansas, Nebraska, Iowa, and South Dakota (Broyles et al., 2004). To evaluate the performance of the proposed approach, coupled RL-OPs, we have chosen Sioux Falls, located in South Dakota, as our study region. The road and power CIs of the study region are respectively represented in Fig. 5a and b (for more details about these CIs refer to He et al., 2016). For performance evaluation, we generated several tornado scenarios for the study region that concurrently cause some disruptions in the power and road CIs.

5.2. Scenario generation

Tornado forecasts and warnings in the U.S. are exclusively issued by the National Weather Service, operating under the National Oceanic and Atmospheric Administration (NOAA). According to NOAA reports, tornadoes exhibit variable movement patterns, although their predominant trajectories are typically from southwest to northeast and from west to east (Roger, 2021). Most tornadoes have durations of <10 min. Utilizing data on tornado path lengths since 1950, the average distance covered by tornadoes is approximately 3.5 miles. This information serves as the basis for generating realistic tornado scenarios.

For scenario generation, we consider four primary movement directions for the tornado: southwest→northeast, west→east, southeast→northwest, and east→west. Additionally, we consider three options for tornado path length (2.5, 3.5, and 4.5 miles) and severity (low, medium, and high). Within the tornado's movement path (yellow line segments in Fig. 5c and d) and its affected region (green areas in Fig. 5c and d), a varying percentage (30 %, 60 %, and 90 %) of links are disrupted at different severity levels. For example, at the low severity level, only 30 % of the links located within the tornado's affected area are randomly selected as disrupted links. This ratio increases to 60 % and 90 % for medium and high severity levels, respectively. This approach enables us to generate small, medium, and large size problem



(a) Individual rewards for the power and road CIs.

(b) Aggregated rewards for the entire community.

Fig. 7. The impact of the number of recovery teams on the nominal rewards of uncoordinated policies.

instances. In Fig. 5c and d, we show a sample tornado trajectory with a southwest→northeast movement direction to provide more details about the scenario generation process. Fig. 5c shows that links (21–24), (21–22), (20–22), (15–19), (19–20), (18–20), and (17–19) in the road CI are located within the tornado's potential damage area. According to Fig. 5d, links (16–18), (15–16), (14–15), (13–14), (14–19), (19–20), and (23–24) in the power CI are located within the tornado's potential damage area. In a scenario with a high severity level, all of these links are considered disrupted in the road and power CIs. In a scenario with a low severity level, three links are randomly selected from these sets as disrupted links for that scenario.

The recovery time for each disrupted link is proportional to its length and is considered a random variable with a uniform distribution, varying 20 % around its average value. Each scenario is evaluated under three different numbers of recovery teams (3, 6, and 9 crews) to assess the approach's performance across varying levels of resource availability. The prerequisite set for each disrupted link in a CI is determined based on spatial closeness. For example, the prerequisite set of a disrupted link in a power CI includes 0, 1, 2, or 3 randomly selected disrupted links in the road CI that are located in a given spatial proximity to that power link.

The learning and exploration parameters used in the original (Eqs. (16) and (25)) and revised (Eqs. (31) and (32)) Bellman's equations for distinct and coupled RL-OPs are as follows: $\alpha = 0.25$, $\gamma = 1$, and $\lambda = 0.1$. The RL-OPs are coded using Python 3.10.0 and PyCharm IDE. The optimization models within the learning environment of RLs are coded using Gurobi Optimizer version 9.5.2. The computer used to run the scenarios is 2.40 GHz Intel Core i9- 10885H CPU with 64 GB of RAM.

5.3. Results

The results of solving the generated scenarios are summarized in Tables 1 and 2. Each scenario is solved by two different approaches: (1) distinct RL-OPs (explained in Sections 4.1 and 4.2), which generate uncoordinated restoration policies for the power and road CIs without considering their interdependencies, and (2) coupled RL-OPs (explained in Section 4.3), which generate coordinated policies that account for interdependencies between the CIs.

The restoration policies generated by these two approaches will be evaluated from two perspectives: (i) the feasibility of the policies: this determines the ratio of policies that can be implemented in practice as they are and can generate nominal rewards predicted by their corresponding approach, and (ii) the quality of the policies: this determines the actual reward provided by the policies to the community. For feasible policies, the nominal and actual rewards are equal. Infeasible policies must be modified according to CI interdependencies, resulting in actual rewards lower than the nominal predicted values.

5.3.1. Restoration policies generated by the distinct RL-OPs

The policies generated by distinct RL-OPs may be infeasible in practice as they are, because the agents (representing CI decision-makers) do not consider CI interdependencies during policy generation. Columns 9 and 10 in Table 1 show the feasibility of policies generated by distinct RL-OPs for the power and road CIs. The actual rewards achieved by infeasible policies (measured by the improvement in the service they provide to the community) differ from the nominal rewards predicted by RL-OPs. To calculate their real reward, we modified the implementation of these policies to account for CI interdependencies. For instance, the implementation of a decision (including a set of selected disrupted links) in a generated policy is delayed until all prerequisite links in interdependent CIs are restored. Meanwhile, restoration priority is given to the next decision in the policy that is implementable due to its prerequisites. If there is no such decision, the teams remain idle for a time unit (e.g., a day) and check the decisions again in the next time unit. After this modification, the actual rewards of the policies (e.g., increased demand fulfillment capability for the power CI and reduced travel time/cost for travelers in the road CI) are recalculated. These actual rewards and their nominal values are summarized in Columns 11–13 and Columns 6–8 of Table 1, respectively. A comparison of actual and nominal rewards highlights the implementation bias resulting from uncoordinated restoration planning for interdependent CIs.

Results summarized in Columns 9 and 10 of Table 1 show that 94.4 percent of policies generated by distinct RL-OPs for the power and road CIs are not feasible and implementable in practice. As in current practices, adopting and implementing these policies will result in some implementation bias.

A comparison of actual and nominal rewards for the power CI in different disaster scenarios reveals that the actual additional service (e.g., extra fulfilled power demand) provided by uncoordinated policies is up to 25.4 % lower than their nominal predicted values (Fig. 6a). This reduction ratio is called implementation bias of uncoordinated restoration for the power CI. Similarly, for the road CI, the actual extra service (e.g., reduction in total traffic time/cost of the road network) provided by uncoordinated policies can be up to 34.3 % less than their nominal predicted values (Fig. 6a). Aggregation of extra services provided by both power and road CIs (equal to the sum of the extra services provided by each CI) shows that the lack of coordination among CI decision-makers in the post-disaster restoration process leads to an implementation bias ranging from 2.3 % to 31.4 % (Fig. 6b). These results are summarized in the following observation:

Observation 1. *The lack of coordination among decision-makers in the post-disaster restoration process of interdependent CIs results in infeasible policies in 94.4 percent of the time. Modification of these policies to make them implementable may lead to up to a 31.4 % reduction in their expected*

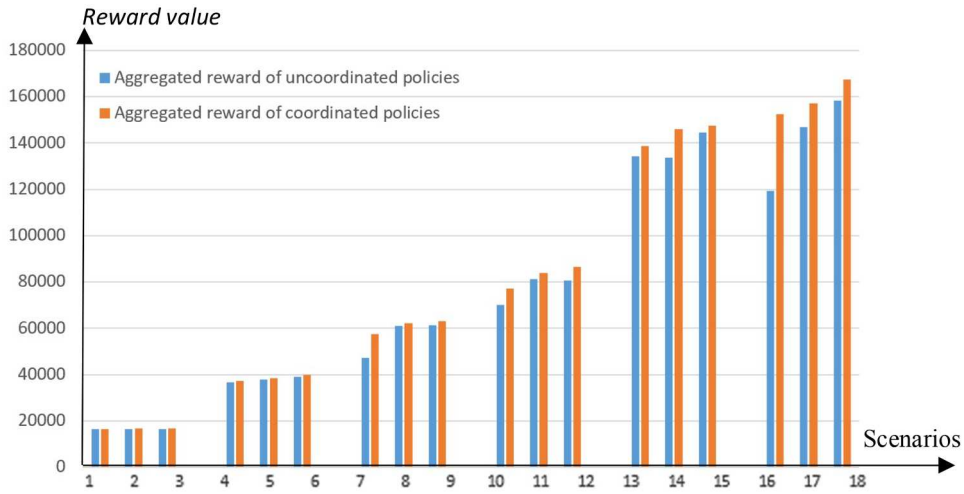


Fig. 8. The comparison of the aggregated rewards generated by the coordinated and uncoordinated policies.

service provision capabilities. These numerical results demonstrate that the lack of coordination among CI decision-makers imposes a substantial burden on the post-disaster resilience of communities facing potential disasters.

5.3.2. Influential factors in the implementation bias of uncoordinated restoration policies

As depicted in Table 1 and Fig. 7, the nominal extra services (called rewards) expected to be provided by the uncoordinated restoration policies increase with the number of recovery teams. This trend is observed across the nominal rewards of the power CI, road CI, and their aggregated rewards. The rationale behind this lies in the dependency of each disrupted link's recovery time on the number of teams assigned to it. By increasing the number of recovery teams, more teams can be assigned to the links selected for restoration. Therefore, less time is needed to restore these links in the CIs, and links become operative more quickly. This swift availability of links enables the CIs to provide better services to the community.

However, a similar trend is not necessarily seen in the actual rewards of uncoordinated policies (refer to Columns 11–13 in Table 1). This happens because the implementation bias amount in each scenario depends on two sets of factors: (1) First Influential Factor Set (FIFS): This includes the number and recovery times of prerequisite links for the disrupted links within that scenario, and (2) Second Influential Factor Set (SIFS): This entails the number of recovery teams available for the CIs within that scenario.

- *Impacts of FIFS*: In scenarios in which the number and recovery times for prerequisite links of important disrupted links (links whose disruption significantly reduce the service provision capability of the CI) are high, we expect a substantial implementation bias (e.g., Scenarios 7 with the aggregated implementation bias of 24.4 % and Scenario 16 with the aggregated implementation bias of 31.4 % in Table 1). This occurs because uncoordinated policies prioritize these important links in the restoration schedule. However, in practice, their restoration and activation in the network is contingent upon the recovery of their prerequisites in other CIs. The high number and extended recovery times of these prerequisites lead to significant restoration delays for these important links in modified policies. This delay significantly increases the implementation bias of these scenarios.
- *Impacts of SIFS*: The restoration times of disrupted links depends on the number of teams assigned to those links. By increasing the number of recovery teams, more teams can be assigned to the links selected for recovery and average restoration times of links reduce. This reduction accelerates the restoration process of CIs. This means delays caused in the absence of prerequisite links will be shorter under uncoordinated policies, leading to a reduction in the implementation bias.

According to this discussion, by increasing the number of recovery teams in a scenario, two outcomes may occur:

- If the increase in the number of recovery teams does not significantly alter the restoration schedule of disrupted links, the impact of SIFS on reducing implementation bias is more substantial than the impact of FIFS on changing it. As a result, the implementation bias reduces by increasing the number of teams. In the tested scenarios, we can see this trend in Scenarios 7–9 and 16–18 (Column 16 in Table 1).
- If the increase in the number of recovery teams significantly alters the restoration schedule of disrupted links, predicting its impact on the implementation bias becomes challenging. If this rescheduling prioritizes links with a high number and recovery times of prerequisites (since uncoordinated policies do not account for prerequisites), it may increase the implementation bias (as explained in the impacts of FIFS). On the other hand, increasing the number of recovery teams may reduce the implementation bias (as explained in the impacts of FIFS). In this case, the tradeoff between these factors determines the change (reduction or increase) in the implementation bias caused by the increased number of teams. In the tested scenarios, we can see this trend in Scenarios 1–3, 4–6, 10–12, and 13–15 (Column 16 in Table 1).

In our numerical results (summarized in Table 1), the average implementation bias in the power CI for scenarios with 3 recovery teams is 11.9 %. This average bias reduces to 5.8 % and 5.7 % for scenarios with 6 and 9 teams, respectively. For the road CI, the average implementation bias for scenarios with 3, 6, and 9 recovery teams is 14.0 %, 6.7 %, and 4.7 %, respectively. The same trend persists when increasing the number of recovery teams involved in the restoration process of both CIs. The average aggregated bias for scenarios with 3, 6, and 9 recovery teams is 14.0 %, 6.6 %, and 5.0 %, respectively. These findings can be summarized as follows:

Observation 2. *The implementation bias caused by the lack of coordination in the restoration process of interdependent CIs is expected to be more significant under the scarcity of recovery resources. This underscores the importance of coordination in disasters with limited recovery resources.*

5.3.3. Comparison of coordinated and uncoordinated restoration policies

In Table 2, Column 9 illustrates the additional power demand that can be met by the power CI over the T horizon when implementing the

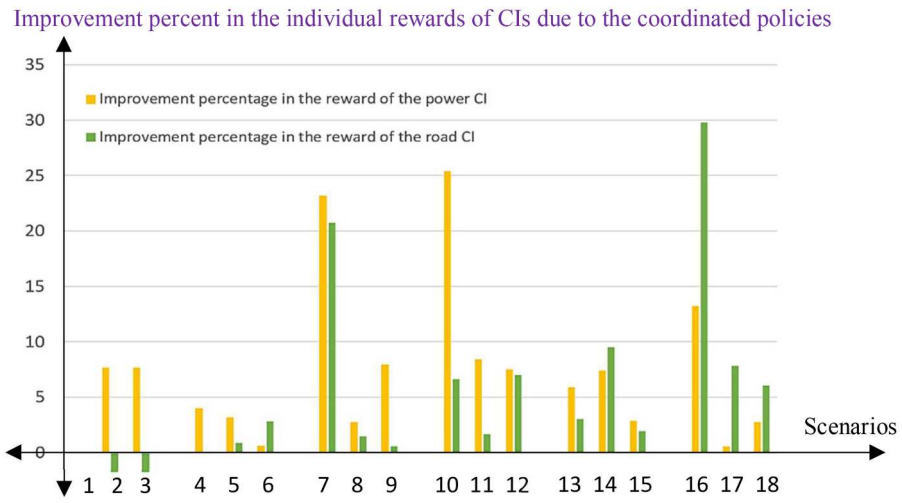


Fig. 9. The improvement percentage made by the coordinated restoration policies in the service provision capabilities of CIs.

coordinated restoration policy generated by the coupled RL-OPs. Column 10 displays the reduction in post-disaster travel time/cost for travelers in the road network over the T horizon when the restoration policy generated by the coupled RL-OPs is applied. Column 11 includes the aggregated rewards of implementing coordinated policies for both CIs. Columns 6, 7, and 8 provide the same information for the uncoordinated policies generated by the distinct RL-OPs. As shown in Columns 12 and 13, all the policies generated by the coupled RL-OPs are feasible and can be implemented without any modifications.

The improvement percentage in the rewards resulting from coordinated policies relative to those generated by uncoordinated policies is displayed in Columns 14, 15, and 16 for the power CI, road CI, and both CIs, respectively. Column 17 shows the computational time for the coupled RL-OPs approach. This time represents the running time required to complete the iterations of the coupled RL-OPs approach. The iterations terminate either when the maximum number of iterations defined for the approach is reached or when the Q values converge. Convergence occurs when the differences between Q values in two successive iterations become smaller than a small, predetermined threshold.

The comparison of aggregated rewards resulting from uncoordinated and coordinated restoration policies reveals that (see Fig. 8):

Observation 3. *The implementation of coordinated policies in the restoration process of interdependent CIs consistently yields higher aggregated service for the community. The overperformance of the coordinated restoration policy can be as high as 27.9 %. The average improvement caused by coordination in scenarios with 3, 6, and 9 recovery teams is 10.7 %, 4.0 %, and 3.4 %, respectively. This implies that the average improvement caused by coordination is more significant in post-disaster circumstances with lower resource availability.*

5.3.4. Importance of coordination in large disaster scenarios with a high number of disruptions

As seen in Fig. 8, the difference between the extra service provision capability of CIs restored using coordinated and uncoordinated policies is more significant in large disaster scenarios with numerous disrupted links (the difference between the orange and blue bars in Fig. 8 grows as the number of disrupted links increases). The improvement due to coordination in large disaster scenarios with 18, 20, 22, and 24 disrupted links (averaged across different numbers of recovery teams) is 8.7 %, 6.7 %, 4.9 %, and 13.5 %, respectively. In contrast, these values drop to 0.8 % and 1.8 % in small scenarios with 6 and 14 disrupted links.

A comparison of improvements due to coordination in the aggregated service/reward of CIs (see Column 16 in Table 2) indicates that the above-mentioned trend is not strict and exhibits some fluctuations. These variations occur because the improvement is scenario-specific and depends on the prerequisites (the number of prerequisite links and their recovery times) of the disrupted links in that scenario. In scenarios in which links with challenging prerequisites (a high number of prerequisites and long recovery times) are disrupted, coordination yields more significant improvements. Overall, the rough trend of improvement due to coordination is an increasing function of scenario size. We can summarize these findings as follows:

Observation 4. *In large disaster scenarios, it is more likely that prerequisites of disrupted links in one CI will be disrupted in its interdependent CI. In such situations, coordinating restoration activities between CIs becomes crucial to ensure that the prerequisites of links selected for restoration in each decision-making stage of a CI have already been recovered in its interdependent CIs. This underscores the critical role of coordination in restoring interdependent CIs during severe disasters that cause extensive link disruption in their networks.*

5.3.5. Impact of coordination on the individual performance of each CI

An individual examination of the improvements made by coordinated restoration policies on the power and road CIs reveals interesting results. In some of the disaster scenarios, the coupled RL-OPs approach attempts to identify and propose feasible/executable restoration policies that are Pareto-optimal compared to the infeasible/un-executable restoration policies generated by distinct RL-OPs. For example, in Scenarios 2 and 3, the reward from the coordinated policy of the coupled RL-OPs is 1.8 % lower for the road CI but 7.7 % higher for the power CI (refer to Columns 14 and 15 in Table 2). However, in other scenarios, the coordinated restoration policies generated by the coupled RL-OPs dominate the uncoordinated policies of distinct RL-OPs (see Fig. 9). For example, in Scenario 16, the rewards from the coordinated restoration policy are 13.2 % higher for the power CI and 29.8 % higher for the road CI. Similarly, in Scenario 13, coordinated policies lead to a 5.9 % higher reward for the power CI and a 3.0 % higher reward for the road CI. These findings can be summarized as follows:

Observation 5. *The proposed coupled RL-OPs approach always generate feasible solutions. These solutions are optimal or at least pareto-optimal in comparison to the infeasible policies generated by distinct RL-OPs.*

6. Closing remark

In this paper, we developed a new approach called coupled RL-OPs, which leverages the decision-making strengths of optimization models through RL. This technique has been used to make restoration decisions

for a set of disrupted interdependent CIs operating in a decentralized context. The proposed technique enables us to make several contributions to the field of CI resilience: (i) It facilitates coordinated restoration for a set of interdependent CIs controlled by separate decision-makers with limited intention for information sharing, (ii) It incorporates uncertain recovery times and dynamic numbers of recovery times in CI restoration planning, and (iii) The approach is flexible enough to make several restoration decisions (e.g., restoration scheduling and resource allocation) simultaneously.

The coupled RL-OPs approach was applied to make restoration decisions for the road and power CIs in Sioux Falls, South Dakota, under several tornado scenarios. Numerical results demonstrate the effectiveness of the coupled RL-OPs in generating superior restoration policies that outperform uncoordinated policies neglecting interdependencies. The ineffectiveness of uncoordinated policies becomes more pronounced in the presence of insufficient restoration resources (e.g., when there are few recovery teams) in most scenarios. On average, the advantage of coordination is expected to be more significant in large-scale disasters with significant disruptions across the interdependent CIs of a community. The proposed approach clearly enhances the post-disaster resilience of communities and cities against disruptive events and disasters affecting their CIs.

The suggested future research directions to expand this study are as follows: (1) Theoretical expansion: The computational time of the coupled RL-OPs increases exponentially with the size of problem instances. Enhancing the computational efficiency of this mechanism is possible by incorporating and integrating deep learning; (2) Application expansion: Beyond disaster management, the proposed coupled RL-OPs can be applied to make synergistic decisions in a wide range of decentralized yet interdependent systems, such as designing risk mitigation policies for interdependent yet autonomous companies within supply chains and devising infection control policies for a set of interdependent wards in healthcare facilities.

CRediT authorship contribution statement

Namrata Saha: Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing. **Shabnam Rezapour:** Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Funding acquisition. **Nazli Ceren Sahin:** Methodology, Writing – review & editing. **M. Hadi Amini:** Conceptualization, Writing – review & editing, Funding acquisition.

Declaration of competing interest

Acknowledgement

This research was financially supported by the National Science Foundation (grant 2108003), the Institute for Resilient and Sustainable Coastal Infrastructure (InteRaCt) and Institute of Environment (FIU preeminent programs), and National Center for Transportation Cybersecurity and Resiliency (TraCR) (a U.S. Department of Transportation National University Transportation Center) headquartered at Clemson University, Clemson, South Carolina, USA.

Data availability

Data will be made available on request.

Acknowledgement

This work was partly supported by the National Science Foundation (Grant 2108003) and FIU preeminent programs of the Institute for Resilient and Sustainable Coastal Infrastructure (InteRaCt), and the Institut of Environment. Also, this work is based upon the work partly supported by the National Center for Transportation Cybersecurity and Resiliency (TraCR) (a U.S. Department of Transportation National University Transportation Center) headquartered at Clemson University, Clemson, South Carolina, USA. Any opinions, findings, conclusions, and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of TraCR, and the U.S. Government assumes no liability for the contents or use thereof.

Appendix A

Table A1
Notation used in the paper.

Sets	
$G^P\left(N^P, L^P\right)$	The power network as a directed graph with a set of nodes, N^P and a set of links, L^P={$l = (n,n')$} where n and $n' \in N^P$
N_S^P	The set of supply nodes in the power network that generate power
N_I^P	The set of intermediate nodes in the power network that transfer power
N_D^P	The set of demand nodes in the power network, representing aggregated households in municipal sites
$N^P = N_S^P \cup N_I^P \cup N_D^P$	The set of all nodes in power network
$G^R\left(N^R, L^R\right)$	The road network as a directed graph with a set of nodes, N^R and a set of links, L^R
N^R	The set of all nodes in the road network, representing urban sites
L^R	The set of all links in the road network, representing roads/highways connecting urban sites
N_O^R	The set of all origin nodes for daily traffics in the road network
N_D^I	The set of all intermediate nodes in the road network
N_D^R	The set of all destination nodes for daily traffics in the road network
OD	The set of all OD pairs in the road network, OD = $\left\{od = (m,m') \middle\vert m \in N_O^R \text{ and } m' \in N_D^R\right\}$
Π_l^P	The set of prerequisite links in the road network that should be restored before restoring link $l \in L^P$ in the power network
Π_l^R	The set of prerequisite links in the power network that should be restored before restoring link $l \in L^R$ in the road network
L'^P	The set of all disrupted links in the power network
$K^P = \{k^P\}$	The set of decision-making stages in the power MDP
$L_k'^P$	A subset of disrupted links in the power network that can be selected for restoration at stage k
S_k^P	The set of potential states, including potential network configurations, in the power network at stage k
$A_{s_k}^P = \left\{a_{s_k}^P\right\}$	The set of actions, including all subsets of links that can be selected for restoration in state s_k, in the power network

(*continued on next page*)

Table A1 (*continued*)

$L^{'R}$	The set of disrupted links in the road network
$K^R = \{k^R\}$	The set of decision-making stages in the road MDP
$L_k^{'R}$	A subset of disrupted links in the road network that can be selected for restoration at stage k
S_k^R	The set of potential states, including potential network configurations, in the road network at stage k
$A_{s_k}^R = \left\{a_{s_k}^R\right\}$	The set of actions, including all subsets of links that can be selected for restoration in state s_k, in the road network
Parameters	
PC_n^P	The daily power generation capacity at supply node $n \in N_S^P$ of the power network
DD_n^P	The daily power demand at demand node $n \in N_D^P$ of the power network
TC_l^P	The flow capacity of link $l \in L^P$in the power network
FC_l^R	The traffic flow capacity of link $l \in L^R$ in the road network
TF_{od}^R	The traffic demand of pair $od \in OD$ in the road network
T	The post-disaster restoration horizon
$\overline{\sigma}_{(n,n')}$	The average restoration time of link $l = (n,n')$
$\sigma_{(n,n')}$	The actual restoration time of link $l = (n,n')$
$\check{\sigma}_{(n,n')}$	The lower bound for the restoration time of link $l = (n,n')$
$\widehat{\sigma}_{(n,n')}$	The upper bound for the restoration time of link $l = (n,n')$
β_l^P	The binary parameter β_l^P is 1 if link $l \in L^P$ is active in the power network; and 0 otherwise
α	The learning convergence speed in RL
γ	The weight of future rewards in RL
β_l^R	The binary parameter β_l^R is 1 if link $l \in L^R$ is active in the road network; and 0 otherwise
tt_l	The travel time/cost for a traveler moving through link $l \in L^R$ in the road network
ct_l^P	The recovery accomplishment time for link $l \in L^P$ in power network
ct_l^R	The recovery accomplishment time for link $l \in L^R$ in power network
C_l	The maximum number of teams that can work concurrently on link l
τ	This index represents RL iterations
M	A big positive value
Λ_k^P	The number of recovery teams available in the power network at stage k
Λ_k^R	The number of recovery teams available in the road network at stage k
λ	The importance of the reward achievable in an interdependent CI for the decision-maker/agent of a CI.
Variables	
$\vartheta'_{L_k^{'P}}$	The actual time needed to recover the links of set $L_k^{'P}$
$\vartheta^*_{L_k^{'P}}$	The minimum time needed to recover the links of set $L_k^{'P}$ if average recovery times occur for the links of this set
$\theta_k^P\left(a_{s_k}^P\right)$	The reward of selecting decision $a_{s_k}^P$ in state s_k in the power MDP
Z^P	The total unfulfilled demand at the demand nodes of the power network during a day
$Q\left(s_k^P, a_{s_k}^P\right)$	The maximum counter-cumulative improvement achievable from stage k to the final stage $\|K\|$, if decision $a_{s_k}^P$ is selected by the power agent in state s_k^P at stage k
π^{P*}	The optimal link restoration policy for the power network
$\vartheta'_{L_k^{'R}}$	The actual time needed to recover the links of set $L_k^{'R}$
$\vartheta^*_{L_k^{'R}}$	The minimum time needed to recover the links of set $L_k^{'R}$ if average recovery times occur for the links of this set
$\theta_k^R\left(a_{s_k}^R\right)$	The reward of selecting decision $a_{s_k}^R$ in state s_k in the road MDP
Z^R	The total post-disaster travel time/cost in the road network
$Q\left(s_k^R, a_{s_k}^R\right)$	The maximum counter-cumulative improvement achievable from stage k to the final stage $\|K\|$, if decision $a_{s_k}^R$ is selected by the road agent in state s_k^R at stage k
π^{R*}	The optimal link restoration policy for the road network
ψ_k^P	The time at which the restoration decisions are made in stage k of the power MDP
ψ_k^R	The time at which the restoration decisions are made in stage k of the road MDP
$\Omega_k^{P,\tau}$	The maximum Q values that are achievable in states of stage k in the power MDP
$\Omega_k^{R,\tau}$	The maximum Q values that are achievable in states of stage k in the road MDP
Decision Variables	
w_l^P	The number of recovery teams assigned to link $l \in L_k^{'P}$ in the power network
UD_n^P	The daily demand that cannot be fulfilled at node $n \in N_D^P$ in the power network
$x^P_{l=(\overrightarrow{n',n})}$	The total power flow from node n' toward node n in the power network
$y^P_{l=(\overrightarrow{n',n})}$	1 if the movement direction of power is from node n' toward node n, and 0 otherwise
b_l^P	The reactance of link $l \in L^P$ in the power network
φ_n	The phase angle of node $n \in N^P$ in the power network
w_l^R	The number of recovery teams assigned to link $l \in L_k^{'R}$ in the road network
$x_l^{R,od}$	The traffic flow through link l related to traffic demand of OD pair od in the road network

Appendix B

Initially, we replace objective function (1) with "Min $\vartheta_{L_k^{'P}}$" and incorporate constraint (B2) into the model:

$$Min\ \vartheta_{L_k^{'P}} \tag{B1}$$

$$\text{S.T. } \vartheta_{L_k^{'P}} \geq \frac{\overline{\sigma}_l}{w^P{}_l}\ \forall l \in L_k^{'P} \tag{B2}$$

$$w^P{}_l \le C_l \ \forall l \in L_k^{'P} \tag{B3}$$

$$\sum_{l \in L_k^{'P}} w^P{}_l \le \Lambda_k^P \tag{B4}$$

$$w^P{}_l \ge 0 \text{ and integer } \forall l \in L_k^{'P} \tag{B5}$$

Next, to linearize the model, we define the set of all possible numbers of teams that can be assigned to link l as $\Phi_l = \{1, 2, 3, \ldots, C_l\}$. We introduce a binary variable v_θ^l, where $v_\theta^l = 1$ if $\theta \in \Phi_l$ crews are assigned to link l, and $v_\theta^l = 0$ otherwise. Consequently, we replace the term $\frac{\overline{\sigma}_l}{w^P{}_l}$ with $\left(\frac{\overline{\sigma}_l}{1v_1^l + 2v_1^l + \ldots + C_l v_{C_l}^l}\right)$ and constraint $\sum_{l \in L_k^{'P}} w^P{}_l \le \Lambda_k^P$ with $\sum_{l \in L_k^{'P}} \sum_{\theta \in \Phi_l} \theta v_\theta^l \le \Lambda_k^P$. Additionally, to ensure that exactly one option is selected as the number of assigned crews to each link, we incorporate constraint (B9) into the model:

$$Min\ \vartheta_{L_k^{'P}} \tag{B6}$$

$$\text{S.T. } \vartheta_{L_k^{'P}} \ge \frac{\overline{\sigma}_l}{1v_1^l + 2v_1^l + \ldots + C_l v_{C_l}^l} \ \forall l \in L_k^{'P} \tag{B7}$$

$$\sum_{l \in L_k^{'P}} \sum_{\theta \in \Phi_l} \theta v_\theta^l \le \Lambda_k^P \tag{B8}$$

$$\sum_{\theta \in \Phi_l} v_\theta^l = 1 \ \forall l \in L_k^{'P} \tag{B9}$$

$$v_\theta^l \in \{0,1\} \ \forall l \in L_k^{'P} \text{ and } \forall \theta \in \Phi_l \tag{B10}$$

We proceed by redefining constraint (B7) as $\left[1\left(\vartheta_{L'k} \times v_1^l\right) + 2\left(\vartheta_{L'k} \times v_2^l\right) + \ldots + C_l\left(\vartheta_{L'k} \times v_{C_l}^l\right)\right] \ge \overline{\sigma}_l$, and replacing $\vartheta_{L'k} \times v_\theta^l$ with $v_\theta^{'l}$. Since $\vartheta_{L'k}$ is continuous and v_θ^l is binary, $v_\theta^{'l}$ must be either 0 or $\vartheta_{L'k}$. To enforce this condition, we introduce three additional constraints, (B13)-(B15), to the model. The resulting linearized model is as follows:

$$\text{Min } \vartheta_{L_k^{'P}} \tag{B11}$$

$$\text{S.T. } 1.v_1^{'l} + 2.v_2^{'l} + \ldots + C_l.v_{C_l}^{'l} \ge \overline{\sigma}_l \ \forall l \in L_k^{'P} \tag{B12}$$

$$v_\theta^{'l} \le M.v_\theta^l \ \forall l \in L_k^{'P} \text{ and } \forall \theta \in \Phi_l \tag{B13}$$

$$v_\theta^{'l} \le \vartheta_{L'k} + M\left(1 - v_\theta^l\right) \ \forall l \in L_k^{'P} \text{ and } \forall \theta \in \Phi_l \tag{B14}$$

$$v_\theta^{'l} \ge \vartheta_{L'k} - M\left(1 - v_\theta^l\right) \ \forall l \in L_k^{'P} \text{ and } \forall \theta \in \Phi_l \tag{B15}$$

$$\sum_{\theta \in \Phi_l} v_\theta^l = 1 \ \forall l \in L_k^{'P} \tag{B16}$$

$$\sum_{l \in L_k^{'P}} \sum_{\theta \in \Phi_l} \theta v_\theta^l \le \Lambda_k^P \tag{B17}$$

$$v_\theta^l \in \{0,1\} \text{ and } v_\theta^{'l} \ge 0 \ \forall l \in L_k^{'P} \text{ and } \forall \theta \in \Phi_l \tag{B18}$$

References

Aksu, D. T., & Ozdamar, L. (2014). A mathematical model for post-disaster road restoration: Enabling accessibility and evacuation. *Transportation Research Part E: Logistics and Transportation Review, 61*, 56–67.

Alkhaleel, B. A., Liao, H., & Sullivan, K. M. (2022). Model and solution method for mean-risk cost-based post-disruption restoration of interdependent critical infrastructure networks. *Computers & Operations Research, 144*, Article 105812.

Almoghathawi, Y., Barker, K., & Albert, L. A. (2019). Resilience-driven restoration model for interdependent infrastructure networks. *Reliability Engineering & System Safety, 185*, 12–23.

Amini, M. H., Parsa Moghaddam, M., & Karabasoglu, O. (2017). Simultaneous allocation of electric vehicles' parking lots and distributed renewable resources in smart power distribution networks. *Sustainable Cities and Society, 28*, 332–342.

Averbakh, I. (2012). Emergency path restoration problems. *Discrete Optimization, 9*(1), 58–64.

Averbakh, I., & Pereira, J. (2012). The flowtime network construction problem. *Iie Transactions, 44*(8), 681–694.

Baxter, M., Elgindy, T., Ernst, A. T., Kalinowski, T., & Savelsbergh, M. W. (2014). Incremental network design with shortest paths. *European Journal of Operational Research, 238*(3), 675–684.

Bhuiyan, T. H., Medal, H. R., & Harun, S. (2020). A stochastic programming model with endogenous and exogenous uncertainty for reliable network design under random disruption. *European Journal of Operational Research, 285*(2), 670–694.

Bienstock, D., & Mattia, S. (2007). Using mixed-integer programming to solve power grid blackout problems. *Discrete Optimization, 4*(1), 115–141.

Botterud, A., Ilic, M. D., & Wangensteen, I. (2005). Optimal investments in power generation under centralized and decentralized decision making. *IEEE Transactions on Power Systems, 20*(1), 254–263.

Broyles, C., Crosbie, C., & Center, S. P. (2004). P5. 6 Evidence of Smaller Tornado Alleys Across the United States Based on a Long Track F3 To F5 Tornado Climatology Study From 1880 To 2003. In *22nd Conf. Sev. Local Storms, Hyannis, Massachusetts, Am. Meteorol. Soc (Vol. 5)*.

Bush, G.W. (2003). Homeland security presidential directive 5. National Security Presidential Directives. https://nrt.org/sites/2/files/HSPD_4.pdf.

Caruson, K., & MacManus, S. A. (2008). Disaster vulnerabilities: How strong a push toward regionalism and intergovernmental cooperation? *The American Review of Public Administration, 38*(3), 286–306.
Cavdaroglu, B., Hammel, E., Mitchell, J. E., Sharkey, T. C., & Wallace, W. A. (2013). Integrating restoration and scheduling decisions for disrupted interdependent infrastructure systems. *Annals of Operations Research, 203*, 279–294.
Çelik, M., Ergun, Ö., & Keskinocak, P. (2015). The post-disaster debris clearance problem under incomplete information. *Operations Research, 63*(1), 65–85.
Chen, J., Thorp, J. S., & Dobson, I. (2005). Cascading dynamics and mitigation assessment in power system disturbances via a hidden failure model. *International Journal of Electrical Power & Energy Systems, 27*(4), 318–326.
Chertkov, M., Backhaus, S., & Lebedev, V. (2015). Cascading of fluctuations in interdependent energy infrastructures: Gas-grid coupling. *Applied Energy, 160*, 541–551.
Ding, Z., Huang, Y., Yuan, H., & Dong, H. (2020). Introduction to reinforcement learning. *Deep reinforcement learning: Fundamentals, research and applications* (pp. 47–123).
Dobson, I., Carreras, B.A., Lynch, V.E., & Newman, D.E. (2001, January). An initial model for complex dynamics in electric power system blackouts. In hicss.
Roger E. (2021). The online Tornado FAQ. Storm Prediction Center, NOAA., Retrieved on January 15, 2024: https://www.spc.noaa.gov/faq/tornado/.
Fakhry, R., Hassini, E., Ezzeldin, M., & El-Dakhakhni, W. (2022). Tri-level mixed-binary linear programming: Solution approaches and application in defending critical infrastructure. *European Journal of Operational Research, 298*(3), 1114–1131.
Fan, J., He, P., Li, C., Zhao, C., & Ji, Y. (2024). A post-disaster restoration model for power-gas-transportation distribution networks considering spatial interdependency and energy hubs. *Electric Power Systems Research, 233*, Article 110505.
Fang, Y. P., & Zio, E. (2019). An adaptive robust framework for the optimization of the resilience of interdependent infrastructures under natural hazards. *European Journal of Operational Research, 276*(3), 1119–1136.
Farzaneh, M. A., Rezapour, S., Baghaian, A., & Amini, M. H. (2023). An integrative framework for coordination of damage assessment, road restoration, and relief distribution in disasters. *Omega, 115*, Article 102748.
Anon. Federal Emergency Management Agency (2018) National Disaster Recovery Framework https://www.fema.gov/national-disaster-recovery-framework (Retrieved on April 3, 2024).
Garay-Sianca, A., & Pinkley, S. G. N. (2021). Interdependent integrated network design and scheduling problems with movement of machines. *European Journal of Operational Research, 289*(1), 297–327.
González, A. D., Dueñas-Osorio, L., Sánchez-Silva, M., & Medaglia, A. L. (2016). The interdependent network design problem for optimal infrastructure system restoration. *Computer-Aided Civil and Infrastructure Engineering, 31*(5), 334–350.
Hafeznia, H., & Stojadinović, B. (2023). ResQ-IOS: An iterative optimization-based simulation framework for quantifying the resilience of interdependent critical infrastructure systems to natural hazards. *Applied Energy, 349*, Article 121558.
He, F., Yin, Y., Wang, J., & Yang, Y. (2016). Sustainability SI: Optimal prices of electricity at public charging stations for plug-in electric vehicles. *Networks and Spatial Economics, 16*, 131–154.
Huang, X., & Wang, N. (2024). Post-disaster restoration planning of interdependent infrastructure Systems: A framework to balance social and economic impacts. *Structural Safety, 107*, Article 102408.
Jasmin, E. A., Ahamed, T. I., & Raj, V. J. (2011). Reinforcement learning approaches to economic dispatch problem. *International Journal of Electrical Power & Energy Systems, 33*(4), 836–845.
Kapucu, N., & Garayev, V. (2013). Designing, managing, and sustaining functionally collaborative emergency management networks. *The American Review of Public Administration, 43*(3), 312–330.
Kaufman, S. M., Qing, C., Levenson, N., & Hanson, M. (2012). Transportation during and after Hurricane Sandy. *Rudin Center for Transportation Policy & Management.* https://rosap.ntl.bts.gov/view/dot/25274.
Kenward, A., Yawitz, D., & Raja, U. (2013). Sewage overflows from hurricane Sandy. *Climate Central.* https://superstormresearchlab.files.wordpress.com/2013/07/climate-central-sewage-overflows-from-hurricane-sandy.pdf.
Leavitt, W. M., & Kiefer, J. J. (2006). Infrastructure interdependency and the creation of a normal disaster: The case of Hurricane Katrina and the city of New Orleans. *Public works management & policy, 10*(4), 306–314.
Lee, E. E., II, Mitchell, J. E., & Wallace, W. A (2007). Restoration of services in interdependent infrastructure systems: A network flows approach. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 37*(6), 1303–1317.
Lee, E. E., Mitchell, J. E., & Wallace, W. A. (2008). Network flow approaches for analyzing and managing disruptions to interdependent infrastructure systems. *Wiley handbook of science and technology for Homeland Security*, 1–9.
Li, S. (2023). *Reinforcement learning for sequential decision and optimal control* (1st ed., pp. 1–460). Singapore: Springer Verlag.
Loggins, R. A., & Wallace, W. A. (2015). Rapid assessment of hurricane damage and disruption to interdependent civil infrastructure systems. *Journal of Infrastructure Systems, 21*(4), Article 04015005.
Maraqa, S. N., Karakoc, D. B., Ghorbani-Renani, N., Barker, K., & González, A. D. (2022). Project schedule compression for the efficient restoration of interdependent infrastructure systems. *Computers & Industrial Engineering, 170*, Article 108342.
Matisziw, T. C., Murray, A. T., & Grubesic, T. H. (2010). Strategic network restoration. *Networks and Spatial Economics, 10*, 345–361.
McDaniels, T., Chang, S., & Reed, D. A. (2008). Characterizing infrastructure failure interdependencies to inform systemic risk. *Wiley Handbook of Science and Technology for Homeland Security*, 1–16.
McDaniels, T., Chang, S., Peterson, K., Mikawoz, J., & Reed, D. (2007). Empirical framework for characterizing infrastructure failure interdependencies. *Journal of Infrastructure Systems, 13*(3), 175–184.
McGuire, M., & Schneck, D. (2010). What if Hurricane Katrina hit in 2020? The need for strategic management of disasters. *Public Administration Review, 70*, s201–s207.
Meyn, S. (2022). *Control systems and reinforcement learning*. Cambridge University Press.
Morales, M. (2020). *Grokking deep reinforcement learning*. Manning Publications.
Moritz, G., Oxenden, M., Albeck-ripka, L., & Ives, M. (2023). At least 32 killed as tornadoes tear through the Midwest and South. *The New York Times*. https://www.nytimes.com/2023/03/31/us/midwest-storms-flood-weather.html.
Morshedlou, N., González, A. D., & Barker, K. (2018). Work crew routing problem for infrastructure network restoration. *Transportation Research Part B: Methodological, 118*, 66–89.
Nilim, A., & El Ghaoui, L. (2005). Robust control of Markov decision processes with uncertain transition matrices. *Operations Research, 53*(5), 780–798.
Nurre, S. G., & Sharkey, T. C. (2014). Integrated network design and scheduling problems with parallel identical machines: Complexity results and dispatching rules. *Networks, 63*(4), 306–326.
Nurre, S. G., Cavdaroglu, B., Mitchell, J. E., Sharkey, T. C., & Wallace, W. A. (2012). Restoring infrastructure systems: An integrated network design and scheduling (INDS) problem. *European Journal of Operational Research, 223*(3), 794–806.
Anon. Office of Electricity Delivery & Energy Reliability, U.S. Department of Energy. (2012). Hurricane Sandy Situation Report. https://www.oe.netl.doe.gov/named_event.aspx?ID=67.
Oruc, B. E., & Kara, B. Y. (2018). Post-disaster assessment routing problem. *Transportation Research Part B: Methodological, 116*, 76–102.
Ouyang, M. (2017). A mathematical framework to optimize resilience of interdependent critical infrastructure systems under spatially localized attacks. *European Journal of Operational Research, 262*(3), 1072–1084.
Paschalidis, I. C., & Kang, S. C. (2008). A robust approach to Markov decision problems with uncertain transition probabilities. *IFAC Proceedings Volumes, 41*(2), 408–413.
Perkins, S. (2002). Tornado alley, USA: New map defines nation's twister risk. *Science News, 161*(19), 296–298.
Pinedo, M. L. (2012). *Scheduling, 29*. New York: Springer.
Rezapour, S., Farahani, R. Z., & Tajik, N. (2021). Impact of timing in post-warning prepositioning decisions on performance measures of disaster management: A real-life application. *European Journal of Operational Research, 293*(1), 312–335.
Rinaldi, S. M., Peerenboom, J. P., & Kelly, T. K. (2001). Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control Systems Magazine, 21*(6), 11–25.
Sahin, H., Kara, B. Y., & Karasan, O. E. (2016). Debris removal during disaster response: A case for Turkey. *Socio-Economic Planning Sciences, 53*, 49–59.
Sang, M., Ding, Y., Bao, M., Li, S., Ye, C., & Fang, Y. (2021). Resilience-based restoration strategy optimization for interdependent gas and power networks. *Applied Energy, 302*, Article 117560.
Sharkey, T. C., Cavdaroglu, B., Nguyen, H., Holman, J., Mitchell, J. E., & Wallace, W. A. (2015). Interdependent network restoration: On the value of information-sharing. *European Journal of Operational Research, 244*(1), 309–321.
Smith, A. M., González, A. D., Dueñas-Osorio, L., & D'Souza, R. M (2020). Interdependent network recovery games. *Risk Analysis, 40*(1), 134–152.
Somers, S., & Svara, J. H. (2009). Assessing and managing environmental risk: Connecting local government management with emergency management. *Public Administration Review, 69*(2), 181–193.
Sun, J., & Zhang, Z. (2020). A post-disaster resource allocation framework for improving resilience of interdependent infrastructure networks. *Transportation Research Part D: Transport and Environment, 85*, Article 102455.
Sütiçen, T. C., Batun, S., & Çelik, M. (2023). Integrated reinforcement and repair of interdependent infrastructure networks under disaster-related uncertainties. *European Journal of Operational Research, 308*(1), 369–384.
Sutton, R. S., & Barto, A. G. (1999). Reinforcement learning. *Journal of Cognitive Neuroscience, 11*(1), 126–134.
Talebiyan, H., & Dueñas-Osorio, L. (2023). Auctions for resource allocation and decentralized restoration of interdependent networks. *Reliability Engineering & System Safety, 237*, Article 109301.
Tzeng, G. H., Cheng, H. J., & Huang, T. D. (2007). Multi-objective optimal planning for designing relief delivery systems. *Transportation Research Part E: Logistics and Transportation Review, 43*(6), 673–686.
Ulusan, A., & Ergun, O. (2018). Restoration of services in disrupted infrastructure systems: A network science approach. *PloS one, 13*(2), Article e0192272.
Anon. US Bureau of Public Roads. Office of Planning. Urban Planning Division. (1964). Traffic assignment manual for application with a large, high-speed computer. US Department of Commerce.
Anon. US Department of Housing and Urban Development, Washington DC. (2013). Hurricane Sandy rebuilding strategy. https://www.hud.gov/sites/documents/HSREBUILDINGSTRATEGY.PDF.
Xu, M., Li, G., & Chen, A. (2024). Resilience-driven post-disaster restoration of interdependent infrastructure systems under different decision-making environments. *Reliability Engineering & System Safety, 241*, Article 109599.