Transportation Research Part E 122 (2019) 1–13

Contents lists available at ScienceDirect

Transportation Research Part E

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

Reliable service systems design under the risk of network access failures

Siyang Xie, Yanfeng Ouyang*

Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

ARTICLE INFO

Keywords:
Facility location
Network access
Service system
Disruptions

ABSTRACT

In service systems with natural or anthropogenic barriers (e.g., rivers, railroads), customers who intend to visit facilities for service must first pass through certain network access points (e.g., bridges, railway crossings). Possible blockage or disruptions of these access points could change the customer-facility assignments or even affect reachability of various facilities, and thus introduce facility reliability and correlation issues. This paper incorporates network access points and their probabilistic failures into a joint optimization framework. A layer of network access points are added and connected to facilities to imply the real-world connections between facilities and access points. The access points are assumed to be subject to disruptions with site-dependent probabilities. We then develop a compact mixed-integer mathematical model to optimize the facility location and customer assignment decisions for the service systems design. Lagrangian relaxation based algorithms are designed to effectively solve the proposed model. Multiple case studies are constructed to test the model and the algorithm, and to demonstrate their performance and applicability.

1. Introduction

In many real-world service systems, customers have to pass through certain network access points to visit facilities for service, and the access points are subject to possible blockage or disruptions. If an access point is blocked/disrupted, some customers may have to change their paths to the assigned facilities, or they may even change their facility choices – some facilities may become unreachable if all associated access points are blocked or disrupted. The unreachability of a facility can be interpreted as the facility's failure to provide service, although the facility itself may always be functioning. For example, in coastal areas or cities where rivers or lakes exist and partition the area into sub-regions, bridges, as the possible only access points to enter/leave the sub-regions, link the partitioned regions as an interconnected network (one distinct example would be the city of Venice, Italy). In such a network, if a bridge is blocked or disrupted due to external factors (e.g., structure or material damages, traffic accidents, congestions), customers originally intend to go through the bridge to visit service facilities like hospitals have to make a detour with a longer transportation distance and/or time, which may lead to significant deterioration of the service quality.

Similar situations could happen in many other contexts. For example, in a ground transportation network consisting of intersecting highway and railway corridors, the highway paths between customers (e.g., residential neighborhoods) and facilities (e.g., fire stations) may cross by railway tracks. If railroad incidents happen and cause railroad blockage, the paths between customers and facilities may be cut off as well. The customers are no longer able to receive service (e.g., emergency response) from its preferable facility in time, which could lead to catastrophic losses (e.g., imagine the fire trucks are blocked at railroad crossings). Recent years

* Corresponding author.
E-mail address: yfouyang@illinois.edu (Y. Ouyang).

https://doi.org/10.1016/j.tre.2018.11.002
Received 23 March 2018; Received in revised form 10 September 2018; Accepted 5 November 2018
Available online 03 December 2018
1366-5545/

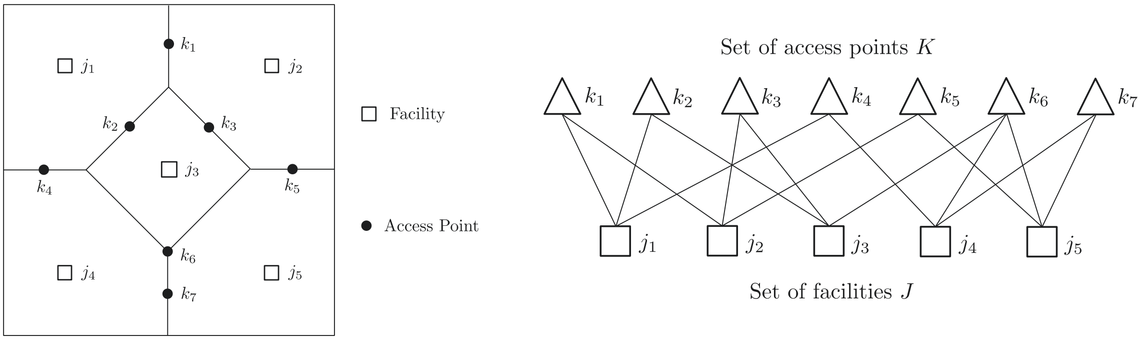
have witnessed a series of rail crashes and derailments that have led to major oil spills, tanker fire or explosions. The train carriages are forced to stop at rail tracks, blocking all crossings to enter or leave the affected region by the explosions. Examples include the recent catastrophic railroad incidents in Casselton, N.D., and in Quebec, Canada (Crummy, 2013; NBC News, 2013). In the U.S. Midwest, the State of Minnesota also expressed strong concerns over the close proximity of hazardous material trains to densely-populated urban areas,[1] and the long blockage of rail crossing by high train traffic volumes which may disrupt emergency response efforts.[2] As a result, the U.S. federal and local regulators have issued a number of orders on railroad incidents prevention and emergency responses deployment so as to enhance rail crossing safety and reliability (Xie et al., 2018b; Gold and Stevens, 2014). This calls for careful design of emergency service facility locations (e.g., fire stations, hospitals) as well as the adjacent network characteristics such that critical resources can be delivered and the customers can be serviced efficiently even under emergency situations.

The planning of such service systems under the risk of network access disruptions involves facility location decisions and customer path design under different realizations of the functioning states of access points. In the literature, there have been plenty of studies employing discrete optimization techniques to determine these reliable planning decisions that minimize the total system costs or maximize the system utility (Drezner, 1995; Daskin, 2011). Most of the existing studies focused on modeling probabilistic disruptions: Snyder and Daskin (2005) formulated stochastic facility location models where facilities are subject to independent disruptions with identical disruption probabilities. Berman et al. (2007) further formulated the reliable facility-location problem as a non-linear mixed-integer mathematical model and designed efficient heuristic solution approaches. Cui et al. (2010) relaxed the assumption of identical probability by allowing the facility disruption probabilities to be site-dependent. Two distinct models (discrete and continuous) and a series of algorithms are proposed. However, these studies assume independent facility disruptions, which are not always realistic in the service systems. Particularly, the “failure” of service facilities may display complex correlation patterns due to their connections with shared access points. The existence of correlations among facilities makes it difficult to even evaluate system performance in a succinct mathematical way; only a few efforts have been made to address correlated facility disruptions (Li et al., 2013; Liberatore et al., 2012; Xie et al., 2015). Specifically, Li et al. (2013) proposed a virtual station structure that transforms a facility network with positively correlated disruptions into an equivalent one with additional virtual supporting stations; these stations are assumed to be subject to independent and identically distributed disruptions. Recently, Xie et al. (2015) presented an overarching recipe of the network transformation theory to address any (positive, negative, or mixed) correlations among facility disruptions. It is proven that a general system of interdependent facilities can be augmented into an equivalent facility-station structure with only independent stations. Xie et al. (2018a) further developed an optimization framework for the reliable facility location planning under generally correlated facility disruptions. The concepts of supporting station structure and quasi-probability are incorporated and customized solution algorithms are designed.

Most existing studies in the literature assume that disruptions to the system occur directly at the facilities themselves. However, in many real-world service systems, the disruptions occur at the network access points, which further make multiple facilities unreachable or even lose functionality. In the literature, the disruptions of network access points are often modeled as network link disruptions (Sullivan et al., 2009; Novak and Sullivan, 2014). These studies, however, focused on developing quantitative measures of network robustness/vulnerability, or applications of equilibrium-based traffic assignments under network disruptions. Facility location design against these types of disruptions (i.e., caused by the failures of network accesses/links), to the best of our knowledge, has not been addressed by these existing models. Those existing studies, either only consider systems with facilities only (Li and Ouyang, 2010; Lu et al., 2015), or involves only i.i.d. disruptions (Berman et al., 2009; Xie and Ouyang, 2016), or lacks methodology to model an additional layer of facilities (i.e., the network access points in our context) (Liberatore et al., 2012; Xie et al., 2015). Therefore, incorporating network access points with site-dependent failures into the service systems design calls for a new methodological framework. In light of these challenges, we extend the concept of station structure in Li et al. (2013) to address the real-world situations with unreliable network access points. Network access points are connected to candidate facility locations based on their real-world relationships/connections. A customer who intends to visit a facility must first pass through one of the facility’s connected access points. The facilities are functioning all the time, but the access points are subject to probabilistic disruptions. Different from Li et al. (2013) that can only handle cases where the station disruption probabilities are all identical, we allow the disruption probabilities of access points to be site-dependent. With the augmented structure with additional access points, we follow the modeling concepts in Cui et al. (2010) to develop a compact mixed-integer mathematical model to address the joint optimization for both facility location and customer access/assignment decisions. Allowing site-dependent disruption probabilities makes our model considerably more difficult than that in Li et al. (2013), as also evidenced in the earlier work in Cui et al. (2010), An et al. (in press). Hence with algorithm ideas similar to those in Cui et al. (2010), we design several customized solution algorithms based on Lagrangian relaxation. Multiple case studies are presented to not only test the model and algorithm but also to draw managerial insights.

The remainder of the paper is organized as follows. Section 2 introduces the incorporation of network access points and presents the mixed-integer mathematical model. Section 3 shows the customized Lagrangian relaxation based solution approaches. In Section 4, several case studies involving various problem and parameter settings are presented. Finally, Section 5 concludes the paper and discusses future research directions.

[1] See http://www.dot.state.mn.us/newsrels/15/03/19oiltrains.html.
[2] See http://www.startribune.com/politics/statelocal/286633141.html.



(a) Facility network with access points (b) Augmented facility-access structure

Fig. 1. Conceptual illustration of the augmented facility-access structure.

2. Model formulation

In this section, we propose a new mixed-integer nonlinear program formulation for the reliable facility location problem under the risk of network access failures.

2.1. Augmented facility-access structure

In some facility systems, customers visit facilities by passing through certain network access points. For example, in Fig. 1(a), to visit facility j_1, a customer may pass through one of the three access points k_1, k_2, k_4 located in the boundaries. We denote K as the set of network access points, and I and J as the sets of customers and candidate facility locations, respectively. As shown in Fig. 1(b), the network access points are connected to facilities to imply the real-world relationships between them, i.e., a facility is connected to an access point if the facility can be reached by customers through the access point. For example, in Fig. 1(b), access point k_1 is connected to facilities j_1 and j_2 as both j_1 and j_2 can be reached by passing through k_1. As such, an access point $k \in K$ could be connected to multiple facilities, and a facility $j \in J$ could be connected to multiple access points as well. The original facility system is consequently augmented into an integrated facility-access structure. We use a binary parameter $l_{kj} = 1$ to indicate that facility j is connected to access point k, or $l_{kj} = 0$ otherwise. Since the network access points are subject to possible disruptions, we further assume that each $k \in K$ is associated with a site-dependent failure probability q_k. [3] The basic mechanism of the augmented structure is defined as follows: a facility remains operational if and only if at least one of its connected access points is functioning (i.e., at least one access point is available to be passed through to reach the facility). Hence, the operating states of the facilities are determined collectively by the states of all access points. For example, in Fig. 1(b), j_1, j_2, j_3 are unreachable while j_4, j_5 are available if and only if access points k_1, k_2, k_3, k_4, k_5, k_6 are all disrupted while k_7 is functioning.

In the facility system, each customer $i \in I$ has a demand μ_i and each candidate location $j \in J$ is associated with a fixed setup cost f_j. Normally, each customer i visits its most preferred facility j for service by passing through access point k (defined as an access-facility pair (k, j) in the rest of the paper). The transportation cost for access-facility pair (k, j) to serve one unit of demand from customer i is denoted as d_{ikj}. [4] Moreover, a penalty cost π_i per unit demand will be incurred if customer i do not receive any service. This situation occurs if no facility is reachable/available, or if the cost of serving customer i by the nearest available facility already exceeds π_i.

Since the access points are possible to be disrupted, each customer may choose multiple access-facility pairs as backups in case her preferable choices are unavailable, whereas each access-facility pair corresponds to one unique backup level. As such, we assume that each customer can select at most a number R of access-facility pairs and pass through the associated access points to visit the corresponding facilities for service, and a customer will pass through its level-r access point if all its level-1, ..., level-$(r-1)$ access points have been disrupted. Note that a customer passes through an access point at no more than one backup level, but may visit a facility through different access points at multiple backup levels. We further add a dummy *emergency* access-facility pair with indices $k = 0, j = 0$ to allow the "penalty assignment", i.e., when a customer losses service. Note that $l_{00} = 1$ and $q_0 = 0$, and the corresponding transportation cost is set to be the penalty cost, i.e., $d_{ikj}|_{k=0,j=0} = \pi_i$, $\forall\, i \in I$. Typically, a customer shall be assigned to the pair $(0, 0)$ at level $R + 1$ and to regular access-facility pairs at backup levels $1, 2, ..., R$. This happens when the penalty cost is

[3] It is worthy noting that Li et al. (2013) and Xie et al. (2015) proposed using homogeneous stations (with equal failure probabilities) to approximate site-dependent failures. However, such approximation procedure will dramatically increase the number of "stations" and the size of the optimization problem. Even for median-size problems in this paper, the number of additional access points needed to approximate the site-dependent failures is very large, and hence the computational burden becomes prohibitive.

[4] In the real world, the distance between a facility and a customer could be affected by disruptions of access points along the shortest path (even when the facility and customer remain accessible). This complicating issue is ignored in this paper for simplicity, but shall be considered in the future.

sufficiently high, and any customer will choose not to give up service as long as there exists some available regular service option (i.e., access-facility pair). When the penalty is not sufficiently large, however, customers may give up service instead of traveling far to obtain regular service. Particularly, if customer i cannot receive service from any access-facility pair (k, j) at a unit cost less than π_i at any backup level $s \in \{1, 2, ..., R\}$, it will choose the emergency pair at level s.

2.2. Formulation

We first define several sets of decision variables. First, variables $\mathbf{X} := \{X_j\}_{j\in J}$ determine the facility locations as follows

$$X_j = \begin{cases} 1 & \text{if a facility is built at } j; \\ 0 & \text{otherwise.} \end{cases}$$

Next, the assignment of customers to access-facility pairs at multiple backup levels is specified by $\mathbf{Y} := \{Y_{ikjr}\}_{i\in I, k\in K\cup\{0\}, j\in J\cup\{0\}, r\in\{1,2,...,R+1\}}$ where

$$Y_{ikjr} = \begin{cases} 1 & \text{if customer } i \text{ visits facility } j \text{ through access point } k \text{ at backup level } r; \\ 0 & \text{otherwise.} \end{cases}$$

Finally, we define $\mathbf{Z} := \{Z_{ikjr}\}_{i\in I, k\in K\cup\{0\}, j\in J\cup\{0\}, r\in\{1,2,...,R+1\}}$, where $Z_{ikjr} \in [0, 1]$ denotes the probability for customer i to visit facility j through access point k at backup level r. The reliable service system design problem (RSSD) under the risk of network access failures is formulated into the following mixed-integer nonlinear program:

$$\text{(RSSD)} \quad \min \quad \sum_{j\in J} f_j X_j + \sum_{i\in I} \sum_{k\in K\cup\{0\}} \sum_{j\in J\cup\{0\}} \sum_{r=1}^{R+1} \mu_i d_{ikj} Z_{ikjr} Y_{ikjr} \tag{1a}$$

$$\text{s.t.} \quad \sum_{r=1}^{R} Y_{ikjr} \leqslant X_j, \quad \forall\, i \in I, k \in K, j \in J, \tag{1b}$$

$$Y_{ikjr} \leqslant l_{kj}, \quad \forall\, i \in I, k \in K \cup \{0\}, j \in J \cup \{0\}, r = 1, 2, ..., R + 1, \tag{1c}$$

$$\sum_{j\in J} \sum_{r=1}^{R} Y_{ikjr} \leqslant 1, \quad \forall\, i \in I, k \in K, \tag{1d}$$

$$\sum_{k\in K} \sum_{j\in J} Y_{ikjr} + \sum_{s=1}^{r} Y_{i00s} = 1, \quad \forall\, i \in I, r = 1, 2, ..., R + 1, \tag{1e}$$

$$\sum_{r=1}^{R+1} Y_{i00r} = 1, \quad \forall\, i \in I, \tag{1f}$$

$$Z_{ikj1} = l_{kj}(1 - q_k), \quad \forall\, i \in I, k \in K \cup \{0\}, j \in J \cup \{0\}, \tag{1g}$$

$$Z_{ikjr} = l_{kj}(1 - q_k) \sum_{k'\in K} \sum_{j'\in J} \frac{q_{k'}}{1 - q_{k'}} Z_{ik'j'(r-1)} Y_{ik'j'(r-1)}, \quad \forall\, i \in I, k \in K \cup \{0\}, j \in J \cup \{0\}, r = 2, 3, ..., R + 1, \tag{1h}$$

$$X_j,\, Y_{ikjr} \in \{0, 1\},\; 0 \leqslant Z_{ikjr} \leqslant 1, \quad \forall\, i \in I, k \in K \cup \{0\}, j \in J \cup \{0\}, r = 1, 2, ..., R + 1. \tag{1i}$$

The objective function (1a) presents the expected system cost as the summation of the fixed facility cost, the expected total transportation cost, and the expected penalty cost. Constraints (1b) and (1c) require that customers can only visit open facilities through access points that are connected to the facilities. Constraints (1d) ensure that for each customer, any access point is selected to be passed through at no more than one backup level. Constraints (1e) enforce that at each level r, any customer $i \in I$ is either assigned to a regular access-facility pair, or assigned to the dummy access-facility pair at an earlier level $s \leqslant r$, while constraints (1f) postulate that each customer is assigned to the dummy access-facility pair at a certain backup level $r \in \{1, 2, ..., R + 1\}$. Constraints (1g) and (1h) recursively define the assignment probabilities Z_{ikjr}: at level $r = 1$, the probability Z_{ikjr} is simply the probability for access k to function; at level $r > 1$, the probability Z_{ikjr} equals $(1 - q_k)q_{k'}/(1 - q_{k'})Z_{ik'j'r-1}$ if that customer i is assigned to j' through k' at level $r - 1$. Constraints (1i) are integrality constraints.

The above formulation is nonlinear because of the nonlinear terms $\{Z_{ikjr}Y_{ikjr}\}$ contained in the objective function and constraints (1h). Therefore, we linearize each $Z_{ikjr}Y_{ikjr}$, which is a product of a bounded continuous variable and a binary variable, by applying the technique introduced by Sherali and Alameddine (1992). Specifically, a new continuous variable W_{ikjr} is introduced to equivalently replace $Z_{ikjr}Y_{ikjr}$ by enforcing four additional sets of constraints (i.e., (2c)–(2f)), and to transform (RSSD) into the following mixed-integer linear program formulation for the linearized reliable service system design (LRSSD):

$$\text{(LRSSD)}\quad \min \quad \sum_{j\in J} f_j X_j + \sum_{i\in I}\sum_{k\in K\cup\{0\}}\sum_{j\in J\cup\{0\}}\sum_{r=1}^{R+1} \mu_i d_{ijk} W_{ikjr} \tag{2a}$$

s.t. (1b)–(1g),

$$Z_{ikjr} = (1-q_k)\cdot \sum_{k'\in K}\sum_{j'\in J} \frac{q_{k'}}{1-q_{k'}} W_{ik'j'(r-1)},$$
$$\forall\, i\in I,\, k\in K\cup\{0\},\, j\in J\cup\{0\},\, r=2,3,\dots,R+1, \tag{2b}$$

$$W_{ikjr} \leqslant Z_{ikjr} + 1 - Y_{ikjr},\quad \forall\, i\in I,\, k\in K\cup\{0\},\, j\in J\cup\{0\},\, r=1,2,\dots,R+1, \tag{2c}$$

$$W_{ikjr} \geqslant Z_{ikjr} + Y_{ikjr} - 1,\quad \forall\, i\in I,\, k\in K\cup\{0\},\, j\in J\cup\{0\},\, r=1,2,\dots,R+1, \tag{2d}$$

$$W_{ikjr} \leqslant Y_{ikjr},\quad \forall\, i\in I,\, k\in K\cup\{0\},\, j\in J\cup\{0\},\, r=1,2,\dots,R+1, \tag{2e}$$

$$W_{ikjr} \geqslant -Y_{ikjr},\quad \forall\, i\in I,\, k\in K\cup\{0\},\, j\in J\cup\{0\},\, r=1,2,\dots,R+1, \tag{2f}$$

$$X_j,\, Y_{ikjr} \in \{0,1\},\quad \forall\, i\in I,\, k\in K\cup\{0\},\, j\in J\cup\{0\},\, r=1,2,\dots,R+1, \tag{2g}$$

$$0 \leqslant W_{ikjr} \leqslant 1,\quad \forall\, i\in I,\, k\in K\cup\{0\},\, j\in J\cup\{0\},\, r=1,2,\dots,R+1. \tag{2h}$$

In theory, this mixed-integer linear program (LRSSD) is compact and polynomial in size (which is much smaller as compared to the exponential size of the scenario-based model), and thus could potentially be solved by commercial solvers such as CPLEX and Gurobi. However, considering the combinatorial nature of the problem as well as the difficulty associated with site-dependent probabilities, existing solvers generally take an excessively long computation time for even moderately sized instances, as we will show with numerical examples in Section 4. Therefore, in the next section, we develop customized solution approaches to efficiently solve (LRSSD).

3. Solution approach

3.1. Lagrangian relaxation

We relax constraints (1b) in (LRSSD) with Lagrangian multipliers $\{\lambda_{ikj}\}_{\forall i\in I, \forall k\in K, \forall j\in J}$ and move them as penalty terms, which yields the following objective function

$$\sum_{j\in J}\left(f_j - \sum_{i\in I}\sum_{k\in K}\lambda_{ikj}\right)X_j + \sum_{i\in I}\sum_{k\in K\cup\{0\}}\sum_{j\in J\cup\{0\}}\sum_{r=1}^{R+1}\mu_i d_{ikj} W_{ikjr} + \sum_{i\in I}\sum_{k\in K}\sum_{j\in J}\lambda_{ikj}\sum_{r=1}^{R} Y_{ikjr}.$$

With this relaxation, the original model (LRSSD) is decomposed into two unrelated parts, involving the location and assignment variables $\mathbf{X}$ and $\mathbf{Y}$, respectively. The part involving $\mathbf{X}$,

$$\min_{X_j\in\{0,1\},\forall j} \sum_{j\in J}\left(f_j - \sum_{i\in I}\sum_{k\in K}\lambda_{ikj}\right)X_j,$$

can be solved easily by simple inspection; i.e., given any $\{\lambda_{ikj}\}$,

$$X_j = \begin{cases} 1 & \text{if } f_j - \sum_{i\in I}\sum_{k\in K}\lambda_{ikj} < 0;\\ 0 & \text{otherwise.}\end{cases}$$

For the part involving $\mathbf{Y}$, we observe that it can be further separated into individual subproblems, one for each customer. The subproblem (RSSD-SP$_i$) with respect to customer $i\in I$ is

$$\text{(RSSD-SP}_i\text{)}\quad \min \quad \sum_{k\in K\cup\{0\}}\sum_{j\in J\cup\{0\}}\sum_{r=1}^{R+1}\mu_i d_{ikj} W_{kjr} + \sum_{k\in K}\sum_{j\in J}\lambda_{kj}\sum_{r=1}^{R} Y_{kjr} \tag{3a}$$

$$\text{s.t.}\quad Y_{kjr} \leqslant l_{kj},\quad \forall\, k\in K\cup\{0\},\, j\in J\cup\{0\},\, r=1,2,\dots,R+1, \tag{3b}$$

$$\sum_{j\in J}\sum_{r=1}^{R} Y_{kjr} \leqslant 1,\quad \forall\, k\in K, \tag{3c}$$

$$\sum_{k\in K}\sum_{j\in J} Y_{kjr} + \sum_{s=1}^{r} Y_{00s} = 1,\quad \forall\, r=1,2,\dots,R+1, \tag{3d}$$

$$\sum_{r=1}^{R+1} Y_{00r} = 1, \tag{3e}$$

$$Z_{kj1} = l_{kj}(1-q_k), \quad \forall\, k \in K \cup \{0\}, j \in J \cup \{0\}, \tag{3f}$$

$$Z_{kjr} = l_{kj}(1-q_k) \sum_{k' \in K} \sum_{j' \in J} \tfrac{q_{k'}}{1-q_{k'}} W_{k'j'(r-1)},$$
$$\forall\, k \in K \cup \{0\}, j \in J \cup \{0\}, r = 2, 3, ..., R+1, \tag{3g}$$

$$\text{(2c)–(2f)}, \tag{3h}$$

$$Y_{kjr} \in \{0, 1\}, \quad \forall\, k \in K \cup \{0\}, j \in J \cup \{0\}, r = 1, 2, ..., R+1, \tag{3i}$$

$$0 \leqslant W_{kjr} \leqslant 1, \quad \forall\, k \in K \cup \{0\}, j \in J \cup \{0\}, r = 1, 2, ..., R+1. \tag{3j}$$

Note that a certain (RSSD-SP$_i$) only involves the subset of (2c)–(2f) that corresponds to customer index i. (RSSD-SP$_i$) does not involve the relationships between customers, and is much smaller in size compared to the original (LRSSD). Hence, it can be more efficiently handled by commercial solvers. However, since in the Lagrangian relaxation framework, each subproblem (RSSD-SP$_i$) will be solved repeatedly across multiple iterations, we may still encounter heavy computational burden if relying on solvers. Thus, Section 3.2 further develops an optional efficient approximate algorithm for the subproblems.

The summation of the optimal objective values from the above two relaxed parts (which involve **X** and **Y** respectively) provides a lower bound to the original problem (LRSSD). Further, based on the solutions to the relaxed subproblems, we use a simple heuristic to perturb them to obtain a feasible solution to the original problem, which provides an upper bound to (LRSSD). Specifically, we fix the optimal facility location decisions from the first relaxed subproblem involving **X**. Then for each customer i, we sort all built and connected access-facility pairs (i.e., $\{(k, j)\text{: } X_j = 1, l_{kj} = 1\}$) in ascending order of d_{ikj}. With the sorted sequence for each customer i, at every level r, we assign i to pair (k, j) with the smallest d_{ikj} as long as i has never been assigned to k at earlier levels $1, 2, ..., r-1$. The following proposition indicates that the feasible solution from this simple approach is likely to be near-optimum.

Proposition 1. *If $R = |K|$, then in any optimal solution $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$, a customer will be assigned to backup access-facility pairs based on the transportation costs; i.e., if $Y_{ik_1j_1r} = 1$ and $Y_{ik_2j_2(r+1)} = 1$ for some i, r, then $d_{ik_1j_1} \leqslant d_{ik_2j_2}$.*

Proof. Suppose, for a contradiction, that $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ is optimal to (RSSD) but there exist $i, (k_1, j_1), (k_2, j_2)$ and r such that $Y_{ik_1j_1r} = 1$, $Y_{ik_2j_2(r+1)} = 1$ and $d_{ik_1j_1} > d_{ik_2j_2}$. We will show that by swapping (k_1, j_1) and (k_2, j_2) the objective of (RSSD) will decrease. We construct a different solution $(\mathbf{X}', \mathbf{Y}', \mathbf{Z}')$ as follows:

(i) $X'_j = X_j$;

(ii) $Y'_{hlms} = \begin{cases} 1, & \text{if } (h, l, m, s) = (i, k_1, j_1, r+1) \text{ or } (i, k_2, j_2, r); \\ 0, & \text{if } (h, l, m, s) = (i, k_1, j_1, r) \text{ or } (i, k_2, j_2, r+1); \\ Y_{hlms}, & \text{otherwise;} \end{cases}$

(iii) $Z'_{hlms} = \begin{cases} \frac{1-q_{k_2}}{1-q_{k_1}} Z_{ik_1j_1r}, & \text{if } (h, l, m, s) = (i, k_2, j_2, r); \\ q_{k_2} Z_{ik_1j_1r}, & \text{if } (h, l, m, s) = (i, k_1, j_1, r+1); \\ 0, & \text{if } (h, l, m, s) = (i, k_1, j_1, r) \text{ or } (i, k_2, j_2, r+1); \\ Z_{hlms}, & \text{otherwise.} \end{cases}$

By construction, $(\mathbf{X}', \mathbf{Y}', \mathbf{Z}')$ is a feasible solution to (RSSD). We use $\Phi(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ to denote the objective value of (RSSD) associated with $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$, it follows that

$$\begin{aligned} \Phi(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) - \Phi(\mathbf{X}', \mathbf{Y}', \mathbf{Z}') &= \mu_i [d_{ik_1j_1} Z_{ik_1j_1r} + d_{ik_2j_2} Z_{ik_2j_2(r+1)} - (d_{ik_1j_1} Z'_{ik_1j_1(r+1)} + d_{ik_2j_2} Z'_{ik_2j_2r})] \\ &= \mu_i \Bigg[d_{ik_1j_1} Z_{ik_1j_1r} + d_{ik_2j_2} \frac{q_{k_1}(1-q_{k_2})}{1-q_{k_1}} Z_{ik_1j_1r} \\ &\qquad - d_{ik_1j_1} q_{k_2} Z_{ik_1j_1r} - d_{ik_2j_2} \frac{1-q_{k_2}}{1-q_{k_1}} Z_{ik_1j_1r} \Bigg] \\ &= \mu_i Z_{ik_1j_1r} (1-q_{k_2})(d_{ik_1j_1} - d_{ik_2j_2}) \end{aligned}$$

Since $\mu_i Z_{ik_1j_1r}(1-q_{k_2}) \geqslant 0$, $d_{ik_1j_1} > d_{ik_2j_2}$, we have $\Phi(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \geqslant \Phi(\mathbf{X}', \mathbf{Y}', \mathbf{Z}')$, which implies that $\Phi(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ is not optimal. This completes the proof. □

Hence, when $R = |K|$, given the facility location decisions, this heuristic yields an optimal customer assignment plan and a tight upper bound. In case $R < |K|$, it can only guarantee a feasible but not necessarily optimal assignment decisions. However, since the probabilities for large back-up levels to occur, i.e., the product of disruption probabilities of multiple access points, are often smaller by orders of magnitudes, the solution given by this sorting/greedy heuristic shall be quite close to the optimal solution.

In the remainder of the Lagrangian solution framework, we use standard subgradient techniques (Fisher, 2004) to update the multipliers λ; i.e.,

$$\lambda_{ikj}^{n+1} = \lambda_{ikj}^{n} + t_j^n\left(\sum_r Y_{ikjr}^n - X_j^n\right), \tag{4}$$

$$t_j^n = \frac{\xi_n(\Phi^* - \Phi_D(\lambda^n))}{\|\sum_r Y_{ikjr}^n - X_j^n\|^2}, \tag{5}$$

where $\{\lambda_{ikj}^n\}$, $\{t_j^n\}$ represent the Lagrangian multipliers and step sizes in the n-th iteration, respectively, and ξ_n is the scalar to control the value of t_j^n. Notation Φ^* and $\Phi_D(\lambda^n)$ are the best upper bound and the current lower bound, respectively. Specifically, the initial value of ξ_n is set to be $0 < \xi_n \leqslant 2$, and is halved whenever $\Phi_D(\lambda^n)$ fails to increase in some fixed number of iterations. In this way, the values of t_j^n satisfy $t_j^n \to 0$ and $\sum_{i=1}^n t_j^i \to \infty$, which is the fundamental theoretical condition for convergence as claimed in Fisher (2004). The subgradient technique here is adopted to help provide better multipliers and facilitate convergence of the algorithm.

Upon completing the Lagrangian relaxation procedure, the above two bounds, especially the lower bound, may still not be close to optimum. If the Lagrangian relaxation algorithm fails to converge to a small enough gap in a certain number of iterations, we embed it into a branch-and-bound (B&B) framework to further reduce the gap. For details, please refer to Cui et al. (2010), Xie et al. (2018a).

3.2. Approximate solution to subproblems

As we stated before, each (RSSD-SP$_i$) is still a combinatorial problem with exponential complexity in the worst case. Furthermore, considering the large number of nodes we need to explore during the branch-and-bound process, even if we solve each subproblem (e.g., using commercial solvers) relatively quickly (e.g., 1-10s), it may take an excessively long time to complete the entire algorithm and find a good near-optimal solution. Therefore, in this section we develop an approximate algorithm which helps quickly find lower bounds to the relaxed subproblems.

Eqs. (3g) show the connections between Z_{kjr} and $Z_{kj(r-1)}$, $Y_{kj(r-1)}$, which brings difficulty in solving subproblem (RSSD-SP$_i$). Therefore, instead of having Z_{kjr} directly in the formulation, we approximate them with fixed numbers. Let $k_1, k_2, \ldots, k_{|K|+1}$ be an ordering of the access points such that $q_{k_1} \leqslant q_{k_2} \leqslant \ldots \leqslant q_{k_{|K|+1}}$, note that $q_0 = 1$ and $k_{|K|+1} = 0$. We define two additional sets of numbers $\{\alpha_{kr}\}_{\forall k\in K, r\in\{1,2,\ldots,R+1\}}$, $\{\beta_r\}_{\forall r\in\{1,2,\ldots,R+1\}}$, such that

$$\alpha_{kr} = (1-q_k)\prod_{m=1}^{r-1} q_{k_m}, \quad \beta_r = \prod_{m=1}^{r-1} q_{k_m}.$$

We next replace Z_{kjr} and Z_{00r} by their estimates α_{kr} and β_r, respectively. The relaxed subproblem (RSSD-SP$_i$) is further relaxed into

$$(\text{RSSD-RSP}_i) \quad \min \sum_{k\in K}\sum_{j\in J}\sum_{r=1}^{R+1}(\mu_i d_{ikj}\alpha_{kr} + \lambda_{kjr})Y_{kjr} + \sum_{r=1}^{R+1}\mu_i d_{i00}\beta_r Y_{00r} \tag{6a}$$

$$\text{s.t.} \quad Y_{kjr} \leqslant l_{kj}, \quad \forall\, k \in K\cup\{0\}, j\in J\cup\{0\}, r = 1, 2, \ldots, R+1, \tag{6b}$$

$$\sum_{j\in J}\sum_{r=1}^{R} Y_{kjr} \leqslant 1, \quad \forall\, k\in K, \tag{6c}$$

$$\sum_{r=1}^{R+1} Y_{00r} = 1, \tag{6d}$$

$$\sum_{k\in K}\sum_{j\in J} Y_{kjr} + \sum_{s=1}^{r} Y_{00s} = 1, \quad \forall\, r = 1, 2, \ldots, R+1, \tag{6e}$$

$$Y_{kjr} \in \{0, 1\}, \quad \forall\, k\in K\cup\{0\}, j \in J\cup\{0\}, r = 1, 2, \ldots, R+1. \tag{6f}$$

We observe that (RSSD-RSP$_i$) is a combinatorial assignment problem, which can be solved by the Hungarian algorithm (as in Cui et al. (2010)). The following proposition states that the solution to (RSSD-RSP$_i$) provides a lower bound to the relaxed subproblem (RSSD-SP$_i$), and hence it can be embedded into the Lagrangian relaxation framework to facilitate computation speed, yet without affecting the validity of the resulting lower and upper bounds.

Proposition 2. *The solution to (RSSD-RSP$_i$) provides a lower bound to the relaxed subproblem (RSSD-SP$_i$).*

Proof. Let $\mathbf{Y}^*$, $\mathbf{Z}^*$ and $\mathbf{W}^*$ be the optimal solution to (RSSD-SP$_i$). (RSSD-RSP$_i$) can be built from (RSSD-SP$_i$) by replacing Z_{kjr}^* and Z_{00r}^* with α_{kr}^* and β_r^*, respectively, and removing constraints (3f)-(3h) and (3j). Since we are relaxing constraints, the solution $\mathbf{Y}^*$, $\mathbf{Z}^*$ and $\mathbf{W}^*$ should still be feasible to (RSSD-RSP$_i$), and based on the construction of α_{kr} and β_r, we know that $\alpha_{kr}Y_{kjr}$ and $\beta_r Y_{00r}$ are lower bounds of W_{kjr} and W_{00r}, respectively. Therefore, the optimal objective value of (RSSD-RSP$_i$) is a lower bound to the optimal objective of (RSSD-SP$_i$). This completes the proof. □

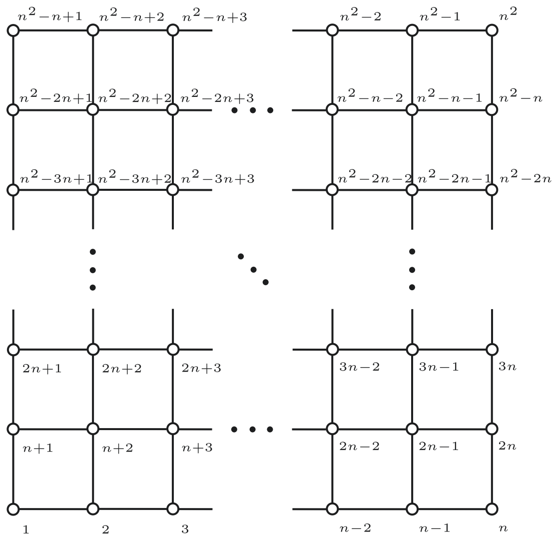


Fig. 2. $n \times n$ hypothetical grid network.

4. Case study

We apply the proposed model and solution algorithms to two examples so as to demonstrate their applicability and performance under different problem and parameter settings. The first example includes a series of hypothetical square grid networks with varying sizes. The second case focuses on planning railroad emergency response facility locations in the Chicago metropolitan area. The main purpose of this example is to illustrate the impacts of various system settings (e.g., heterogeneity) on the optimal design.

The proposed solution algorithms are programmed in C++ and run on a 64-bit Intel i7-3770 computer with 3.40 GHz CPU and 8G RAM. The mixed-integer linear programs (LRSSD) and (RSSD-SP$_i$), if solved directly, are tackled by commercial solver CPLEX 12.4 using up to 4 threads. The reformulated problem (RSSD-RSP$_i$) is solved by the Hungarian algorithm.

4.1. Hypothetical grid networks

For $n \in \{4, 5, 6, 7, 8, 10\}$, an $n \times n$ square grid network is generated to represent a hypothetical study region (e.g., a city like Venice) with n^2 cells (e.g., islands) and $2n(n-1)$ blockage segments (e.g., canal branches), as shown in Fig. 2. We assume that the network corresponds to a coordinate system $(n_1, n_2) \in [1, 2, ..., n] \times [1, 2, ..., n]$, where n_1 increases from left to right, and n_2 increases from bottom to top; the bottom-left and top-right cells have coordinates $(1, 1)$ and (n, n), respectively. We further label the cell at location (n_1, n_2) with index $n_1 + n \times (n_2 - 1)$. The edge length between any two adjacent cells is set to 1, and each cell is considered to be both an individual customer and a candidate facility location. For cell $i = (n_1, n_2)$, the demand is $\bar{\mu}\left(1 + \tau_\mu \cos\left(\pi \frac{n_1 - 1}{n-1}\right)\right)$ and the fixed facility cost is $\bar{f}\left(1 + \tau_f \cos\left(\pi \frac{n_2 - 1}{n-1}\right)\right)$. The values of parameters τ_μ and τ_f determine the extent of heterogeneity of demand and facility cost distribution over the network, such that the customer demand density varies from $\bar{\mu}(1 + |\tau_\mu|)$ on the left side to $\bar{\mu}(1 - |\tau_\mu|)$ on the right side, and the facility cost varies from $\bar{f}(1 + |\tau_f|)$ on the bottom to $\bar{f}(1 - |\tau_f|)$ on the top. In these hypothetical test cases, we set their values to be $\bar{\mu} = 10.0, \bar{f} = 100.0, \tau_\mu = \tau_f = 0.25$, and set the penalty values to be $\pi_i = 1000, \forall\, i \in I$. The middle point of each edge represents the access point (e.g., bridge) through which customers may travel to service facilities. The site-dependent failure probability of the edge between cells i and j are "randomly" generated as $0.015 + 0.005(\text{mod}(i + j, 5) + 1)$. Note that the parameters that describe demand, costs, and probabilities are generated hypothetically to introduce spatial nonlinearity and heterogeneity to the numerical cases; the purpose is to demonstrate the capability of the proposed model. The maximum assignment level is $R = 3$ for all cases.

To solve the reliable facility location problems in these networks, we use three solution approaches: (i) CPLEX directly applied to the linearized original problem (LRSSD); (ii) Lagrangian relaxation based algorithm embedded in a branch-and-bound framework with each subproblem (RSSD-SP$_i$) solved by CPLEX (LR + B&B + CPLEX); and (iii) Lagrangian relaxation based algorithm embedded in a branch-and-bound framework with each subproblem (RSSD-RSP$_i$) solved by the approximate algorithm (LR + B&B + Approx.). The solution time limit is set to be 3600 s. Table 1 summarizes and compares the results obtained by the three approaches for a range of test instances.

Overall, it can be observed that the solution time and solution quality deteriorate with the network size, owing probably to the significant increase in the number of integer variables $\mathbf{Y}$. CPLEX cannot close the optimality gaps for the first three cases despite the relatively small network sizes. For the three larger networks, CPLEX ran out of memory and failed to provide even a feasible solution. The LR + B&B + CPLEX approach can provide a feasible solution within one hour for the first 4 cases, however, the optimality gaps are relatively large except for the smallest 4×4 network. This is because when network size is large, it takes CPLEX a long time to solve even one instance of subproblem (RSSD-SP$_i$), and thus the overall algorithm can only branch on a very limited number of nodes within the time limit. For the 10×10 network, the LR + B&B + CPLEX approach failed to give a feasible solution. In contrast, the LR + B&B

Table 1
Algorithm performance comparison for cases with hypothetical grid networks.

	Network size	Number of facilities	Opt. facility locations	Final UB	Final LB	Final gap (%)	CPU time (s)
CPLEX	4 × 4	1	10	402.09	398.89	0.80	3600
	5 × 5	3	7, 20, 22	635.32	629.22	0.96	3600
	6 × 6	3	8, 23, 26	889.87	880.80	1.02	3600
	7 × 7	–	–	–	–	fail	3600
	8 × 8	–	–	–	–	fail	3600
	10 × 10	–	–	–	–	fail	3600
LR + B&B + CPLEX	4 × 4	1	10	402.09	402.09	0.0	2371
	5 × 5	2	7, 18	638.24	494.32	22.55	3600
	6 × 6	4	3, 23, 25, 34	934.74	598.01	36.02	3600
	7 × 7	5	2, 13, 32, 36, 42	1282.54	728.36	43.21	3600
	8 × 8	–	–	–	–	fail	3600
	10 × 10	–	–	–	–	fail	3600
LR + B&B + Approx.	4 × 4	1	10	402.09	402.09	0.0	1
	5 × 5	3	7, 20, 22	635.32	635.32	0.0	16
	6 × 6	3	8, 23, 26	889.87	889.87	0.0	387
	7 × 7	4	13, 16, 37, 40	1211.59	1211.59	0.0	3244
	8 × 8	6	4, 23, 26, 44, 50, 54	1597.70	1473.76	7.76	3600
	10 × 10	9	17, 22, 45, 59, 61, 77, 82, 89, 94	2520.94	2132.61	15.40	3600
	10 × 10	9	12, 16, 39, 42, 55, 67, 72, 89, 93	2502.48	2176.83	13.01	43200

Table 2
Sensitivity analysis on the penalty value π_i.

Network size	π_i	Opt. facility locations	Final objective	Penalty cost	Final gap (%)	CPU time (s)
6 × 6	1000	8, 23, 26	889.870	4.844	0.0	387
	100	8, 23, 26	885.356	0.599	0.0	420
	10	8, 23, 26	884.799	0.063	0.0	456
	5	8, 23, 26	884.767	0.081	0.0	505
	4	8, 23, 26	884.751	30.247	0.0	458
	3	8, 26, 29	872.070	30.247	0.0	271
	2	∅	720.000	720.000	0.0	2
	1	∅	360.000	360.000	0.0	1

+ Approx. solution approach can obtain exact optimal solutions within 1 h for the first 4 cases. For $n \in \{8, 10\}$, the optimality gaps after 1 h of computation are 7.76%, and 15.40%, respectively. As such, the proposed LR + B&B + Approx. approach clearly outperforms the other two methods in terms of both solution quality and computation time.

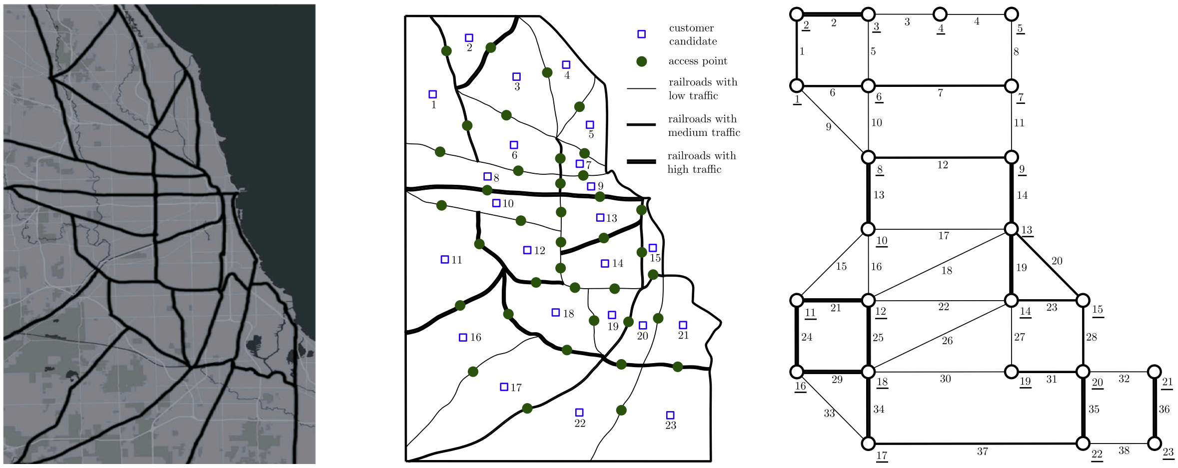
Table 2 presents the sensitivity analysis result on parameter penalty value π_i for the 6 × 6 hypothetical network. Here we assume identical penalty value for all customers, and vary it from 1000 to 1. The results show that when the penalty value is large (e.g., larger than 5), the performance of the algorithm and the solution quality almost remain the same, i.e., we are able to obtain the optimal solution in a few minutes, and the optimal facility locations do not change. On the other hand, when the penalty value is small, the optimal objective value changes significantly because some of the customers may choose to give up service at earlier levels with higher probabilities. This is further demonstrated by comparing cases with penalty values $\pi_i = 10, 5, 4, 2$, i.e., the total penalty cost increases when the value of π_i reduces. In addition, the optimal facility locations change as well. All these are consistent with our discussions before and thus demonstrate the effectiveness of our methodology.

To further check the quality of the solutions and explore the effectiveness of our solution approach, we run the case with network size 10 × 10 for another 11 h. The final gap reduces from 15.49% (after 1 h) to 13.07% (after 12 h), and the objective value of the best known solution decreases slightly by 0.7%, from 2524.32 to 2506.72. We suspect that the best known feasible solution after 1 h of computation is already of a reasonably good quality.

4.2. Railroad emergency response

We now consider the Chicago area, a region with strong railroad network presence as shown in Fig. 3(a). Target areas (e.g., towns and districts) are partitioned and surrounded by railroad segments. We assume there is one major access point (at-grade crossing)[5] on

[5] Each segment may actually include multiple access points. Since a segment cannot be passed through if and only if all access points on it are disrupted, we can approximately consolidate these access points into one "representative" with the "composite" disruption probability $q = \prod_{k \in L} q_k$, where L is the set of all actual access points on this segment. Note that when a segment is long, the distances of passing through these different access points will likely be different. This issue is not addressed in this paper but deserves further study.



(a) The Chicago map (b) The railroad network (c) The abstract representation

Fig. 3. The railroads network setup in Chicago area.

Table 3
Algorithm performance comparison for the railroad emergency response cases.

Index	Failure Prob.		R	Opt. facility	Root	Root	Final	Overall Cost	Trans. cost	Final	Time
	$\bar{q}$	$\hat{q}$		locations	UB	gap (%)	Objective	difference	difference	gap (%)	(s)
1	0.2	0.2	4	5, 6, 10, 14, 22	61376.0	46.0	60282.3	0.81%	3.57%	0	7054
2	0.2	0.2	3	5, 6, 10, 14, 22	61594.4	46.2	60881.2	1.33%	4.50%	2.91	7200
3	0.2	0.2	2	5, 6, 12, 14, 20	64719.2	48.0	62986.0	8.91%	33.03%	6.56	6936
4	0.2	0.1	4	5, 6, 10, 14, 22	61906.3	45.9	61377.6	0.16%	2.26%	0	7200
5	0.2	0.1	3	5, 6, 10, 14, 22	64663.3	48.2	62629.9	0.51%	2.83%	2.50	7200
6	0.2	0.1	2	5, 6, 12, 19, 20	70293.7	51.8	70115.3	3.54%	18.76%	13.07	7200
7	0.2	0	4	3, 7, 10, 14, 22	62690.3	46.5	62355.2	–	–	4.84	7200
8	0.2	0	3	3, 7, 10, 14, 22	65100.8	48.5	64454.7	–	–	8.40	7200
9	0.2	0	2	5, 8, 14, 22	76727.2	55.8	75635.2	–	–	22.35	7200
10	0	0	1	6, 14, 21	57120.0	36.2	57120.0	–	–	0	89

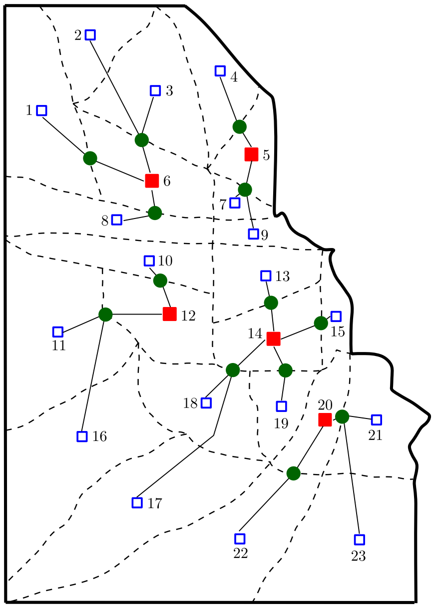
each railroad segment that allow first-responders to reach the regions. A limited number of emergency resource facilities are to be deployed among these regions in anticipation of random emergencies (e.g., fire, incidents). However, the rail crossings are subject to blockages, and hence the emergency resources of a facility might not be accessible if all of its surrounding rail crossings are blocked. The railroad network in Chicago contains $|J| = 23$ candidate facility locations, $|I| = 23$ customers (e.g., cities and towns), and $|K| = 38$ railway-highway crossings. In Fig. 3(b), each line represents a railway segment, each dot represents a railway crossing (access point), and each square (surrounded by multiple railway segments) represents a target demand area as well as a candidate facility location. The demand and fixed cost of each area are set to be proportional to its population and housing price, respectively. The railway crossings serve as access points to built facilities, and they are categorized into three groups, each having a high, median, or low risk of being blocked based on their annual train traffic volumes (denoted by $\mathscr{K}_h$, $\mathscr{K}_m$, $\mathscr{K}_l$, respectively). We further assume that blockage probability of each group can be specified as follows:

$$q_k = \begin{cases} \bar{q} + \hat{q} & \text{if } k \in \mathscr{K}_h \\ \bar{q} & \text{if } k \in \mathscr{K}_m \\ \bar{q} - \hat{q} & \text{if } k \in \mathscr{K}_l \end{cases}$$

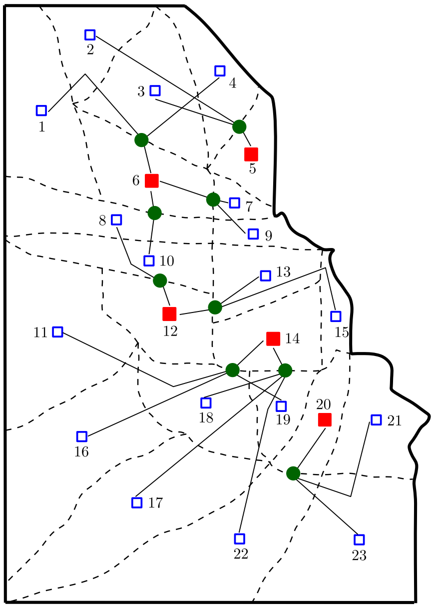
where $\bar{q}$ is the average probability and $\hat{q}$ marks the level of spatial variation. A simplified graph of the network is shown in Fig. 3(c), where each node is a candidate facility location and each link is a railway crossing. The distance d_{ikj} is calculated as the shortest path distance between node i and j through link k based on Fig. 3(d). We assume that a customer receiving service from a facility elsewhere must pass through one of the railway crossings surrounding the facility.

We test our model with a range of $\bar{q}$, $\hat{q}$, and R, so as to examine their impacts on the optimal facility location design and algorithm performances. Case 10 (i.e., $\bar{q} = \hat{q} = 0$) represents the degenerated situation where crossings never get blocked, and hence backup assignments are not necessary (i.e. $R = 1$). For other cases, we assume that $\bar{q} = 0.2$, and $\hat{q} \in \{0, 0.1, 0.2\}$ for identical probability, slight site-dependent probabilities, and high site-dependent probabilities, respectively. The value of R varies from 2 to 4.

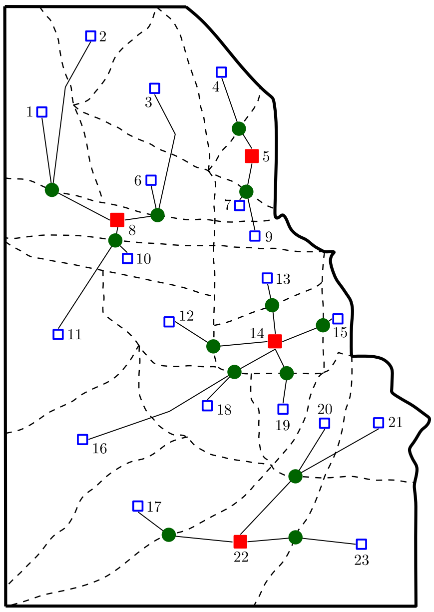
Solutions from our approximate algorithm (LR + B&B + Approx.) are presented in Table 3. The relatively large values of access



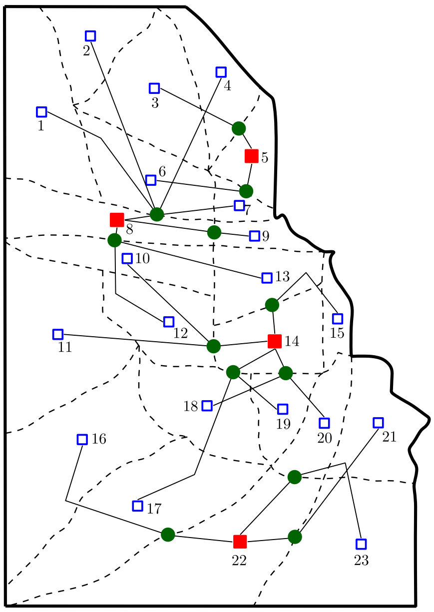
(a) 1st level assignment for case 3



(b) 2nd level assignment for case 3



(c) 1st level assignment for case 9



(d) 2nd level assignment for case 9

Fig. 4. Facility locations and customer assignments at different backup levels of cases 3 and 9.

point failure probabilities have led to longer computation times, i.e., not all cases can be solved to optimality within 2 h. As R increases, the total cost decreases, due to a slightly lower likelihood for the customers to receive the penalty of losing service. In addition, the value of R does have observable impacts on the computation time and the optimal facility location design. These observations are consistent with those in earlier studies by Cui et al. (2010), Li and Ouyang (2012). Existence of access point failures generally has a noticeable impact on the optimal facility locations, total cost (including transportation cost), and the required computation time, if we use the no-failure counterpart (case 10) as the benchmark. Moreover, the spatial heterogeneity (as reflected by the value of $\hat{q}$) is possible to affect the optimal design, e.g., solutions to case 3 (with $\hat{q} = 0.2$), case 6 (with $\hat{q} = 0.1$), and case 9 (with $\hat{q} = 0$) are all quite different. It is worth noting that failing to consider site-dependent probabilities may lead to a cost increase, especially for transportation. For example, if we hold the failure probabilities of case 3 as the ground truth (where $\hat{q} = 0.2$), but solve the problem as if $\hat{q} = 0$. The corresponding solution (the one for case 9) will yield an actual total cost of 68596.4 and a transportation cost of 45636.4 under the assumed ground truth, which are 8.91% and 33.03% larger than the corresponding total cost of 62896.0 and transportation cost of 34566.0 obtained for case 3 (with $\hat{q} = 0.2$), respectively. It shall be also noted, however, that for many other cases, the cost "error" from ignoring access point failure heterogeneity is not as high as those observed in other studies (which directly consider facility failure heterogeneity). This result is somewhat intuitive because the presence of shared access points among the facilities tends to serve as another layer of "buffer" that averages out the spatial heterogeneity. In addition, although the location decisions from our model are not proved to be optimal (i.e., there is still an optimality gap), we believe that the cost differences with truly optimal solutions will remain sufficiently small and reveal similar insights.

Fig. 4 presents the location decisions and assignment path of each customer to access the facilities at each backup level (i.e., 1st and 2nd) for cases 3 and 9 (with $R = 2$). Generally, five facilities {5, 6, 12, 14, 20} are built in case 3, as marked by the solid red squares, while another 4 facilities {5, 8, 14, 22} are built in case 9. For both cases, the built facilities are located at regions with a higher concentration of demands. Specifically, in the southern half of the metropolitan area, due to low demand, only one candidate location (e.g., 20 in case 3 and 22 in case 9) is selected. In contrast, the densely populated northern half always has two or even more built facilities. Moreover, regions/nodes with more access points (e.g., location 14) are more likely to be chosen since they can provide more backup access points/opportunities. As for the customer assignments, the 1st-level assignments can be segregated into multiple groups, with each group clustered around one facility, while the 2nd-level assignments are more intertwined with each other. In addition, a customer may visit the same facility or two different facilities through two different crossings at the two backup levels. For example, in the solution for case 3, customer 4 is first assigned to facility 5 through crossing 4 at level 1, then to facility 6 through crossing 5 at level 2; while customer 1 is first assigned to facility 6 through crossing 6 at level 1, then to facility 6 again through crossing 5 at level 2.

5. Conclusion

In this paper, we study the reliable service systems design problem with considerations of possible network access failures. In service systems where customers pass through certain network access points to visit facilities for service, failures of the network access points could potentially affect the functionality of service facilities, and consequently introduce reliability and correlation issues to the system. We add a layer of network access points, and connect them to facilities to indicate their real-world relationships. With the additional layer of access points, the original facility system is augmented into an integrated facility-access structure, based on which we develop a compact mixed-integer mathematical model to formulate the reliable system design problem. A customized Lagrangian relaxation based algorithm is designed, and multiple case studies are conducted to test the applicability and performance of the proposed model and algorithm. Numerical results provide a range of managerial insights; e.g., site-dependent access point failure probabilities generally lead to higher levels of facility concentration so as to provide more back-up options to the customers.

The study can be further extended in several directions. First, in many real-world contexts, the reachability of facilities or the distance metric could be affected by the network access point disruptions as well. For example, debris from earthquakes or floods may block nearby roadway segments and change the shortest paths between points. Such complicating issues should be addressed in future studies. Second, it will be interesting to consider correlated disruptions of access points, which might be more realistic in many real world contexts (e.g., correlated bridge failures or roadway blockages due to shared hazards). Third, more sophisticated algorithms should be developed to solve larger instances of the proposed problem more efficiently. The Chicago railroad network case study in this paper still requires a long time to reach a relatively small gap, which obviously calls for improvement on the speed and efficiency of the solution algorithm. Additionally, it will also be interesting to apply our methodology to more real-world cases, so as to help policy makers develop engineering and planning guidelines (e.g., on positioning and utilizing emergency response resources) that will lead to more reliable and resilient systems.

Acknowledgment

This research was financially supported in part by the U.S. National Science Foundation through Grants CMMI #1234085, CMMI #1662825, and by the Roadway Safety Institute, the University Transportation Center for USDOT Region 5, which is funded by the United States Department of Transportation's Office of the Assistant Secretary for Research and Technology (OST–R).

Appendix A. Notation list

I	Set of customers;
J	Set of candidate facility locations;
K	Set of network access points;
R	Maximum number of regular backup access-facility pairs;
l_{kj}	Binary parameter indicating whether facility j is connected to access point k;
μ_i	Demand of customer i;
π_i	Penalty cost for customer i to lose service for unit demand
f_j	Fixed setup cost at location j;
q_k	Failure probability of access point k;
d_{ikj}	Transportation cost for access-facility pair (k, j) to serve one unit of demand from customer i;
X_j	Decision variable indicating whether a facility is built at location j;

Y_{ikjr} Decision variable indicating whether customer i visits facility j through access point k at backup level r;
Z_{ikjr} Decision variable indicating the probability for customer i visits facility j through access point k at backup level r;
W_{ikjr} $W_{ikjr} = Z_{ikjr} Y_{ikjr}$;
λ_{ikj}^{n} Lagrangian multiplier for a certain combination of (i, k, j) in the nth iteration;
t_j^n Step size for a certain j in the nth iteration;
ξ_n A scalar to control step size in the nth iteration;
Φ^* Best upper bound in the Lagrangian relaxation algorithm procedure
$\Phi_D(\lambda^n)$ The lower bound obtained in the nth iteration;
α_{kr} The approximation of Z_{kjr};
β_r The approximation of Z_{00r};

References

An, K., Xie, S., Ouyang, Y., 2017;al., in press. Reliable sensor location for object positioning and surveillance via trilateration. Transp. Res. Part B: Methodol (in press).
Berman, O., Krass, D., Menezes, M.B.C., 2007. Facility reliability issues in network p-median problems: strategic centralization and co-location effects. Oper. Res. 55 (2), 332–350.
Berman, O., Krass, D., Menezes, M.B.C., 2009. Locating facilities in the presence of disruptions and incomplete information. Decision Sci. 40 (4), 845–868.
Crummy, B., 2013. Mile-long train carrying crude oil derails, explodes in North Dakota. < http://usnews.nbcnews.com/_news/2013/12/30/22113442-mile-long-train-carrying-crude-oil-derails-explodes-in-north-dakota?lite > (accessed: 2014-03-17).
Cui, T., Ouyang, Y., Shen, Z.-J.M., 2010. Reliable facility location design under the risk of disruptions. Oper. Res. 58 (4), 998–1011.
Daskin, M.S., 2011. Network and Discrete Location: Models, Algorithms, and Applications, second ed. John Wiley & Sons.
Drezner, Z., 1995. Facility Location: A Survey of Applications and Methods. Springer Verlag.
Fisher, M.L., 2004. The lagrangian relaxation method for solving integer programming problems. Manage. Sci. 50 (12_supplement), 1861–1871.
Gold, R., Stevens, L., 2014. U.S. issues emergency testing order to crude oil rail shippers: Move by transportation department response to crude-by-rail accidents. < http://on.wsj.com/1WQaKvt > (accessed: 2014-03-17).
Li, X., Ouyang, Y., 2010. A continuum-approximation approach to reliable facility location design under correlated probabilistic failures. Transp. Res. Part B 44 (4), 535–548.
Li, X., Ouyang, Y., 2012. Reliable traffic sensor deployment under probabilistic disruptions and generalized surveillance effectiveness measures. Oper. Res. 60 (5), 1183–1198.
Li, X., Ouyang, Y., Peng, F., 2013. A supporting station model for reliable infrastructure location design under interdependent disruptions. Transp. Res. Part E 60 (0), 80–93.
Liberatore, F., Scaparra, M.P., Daskin, M.S., 2012. Hedging against disruptions with ripple effects in location analysis. Omega 40 (1), 21–30.
Lu, M., Ran, L., Shen, Z.-J.M., 2015. Reliable facility location design under uncertain correlated disruptions. Manuf. Serv. Oper. Manage. 17 (4), 445–455.
NBC News, 2013. Danger on the tracks: Unsafe rail cars carry oil through us towns. < http://www.nbcnews.com/news/other/danger-tracks-unsafe-rail-cars-carry-oil-through-us-towns-f8C11082948 > (accessed: 2014-03-17).
Novak, D.C., Sullivan, J.L., 2014. A link-focused methodology for evaluating accessibility to emergency services. Decis. Support Syst. 57, 309–319.
Sherali, H.D., Alameddine, A., 1992. A new reformulation-linearization technique for bilinear programming problems. J. Global Optim. 2 (4), 379–410.
Snyder, L.V., Daskin, M.S., 2005. Reliability models for facility location: the expected failure cost case. Transp. Sci. 39 (3), 400–416.
Sullivan, J., Aultman-Hall, L., Novak, D., 2009. A review of current practice in network disruption analysis and an assessment of the ability to account for isolating links in transportation networks. Transp. Lett. 1 (4), 271–280.
Xie, S., An, K., Ouyang, Y., 2018a. Planning facility location under generally correlated facility disruptions: use of supporting stations and quasi-probabilities. Transp. Res. Part B: Methodol (submitted for publication).
Xie, S., An, K., Ouyang, Y., 2018b. Positioning, Planning and Operation of Emergency Response Resources and Coordination Between Jurisdictions. Tech. rep. Center for Transportation Studies, University of Minnesota.
Xie, S., Li, X., Ouyang, Y., 2015. Decomposition of general facility disruption correlations via augmentation of virtual supporting stations. Transp. Res. Part B 80, 64–81.
Xie, S., Ouyang, Y., 2016. Railroad caller districting with reliability, contiguity, balance, and compactness considerations. Transp. Res. Part C: Emerg. Technol. 73, 65–76.