Optimizing System Resilience: A Facility Protection Model With Recovery Time

Chaya Losada
Kent Business School

Maria Paola Scaparra
Kent Business School

Jesse R O’Hanley
Kent Business School

Working Paper Series
Optimizing system resilience: a facility protection model with recovery time

Chaya Losada∗ M. Paola Scaparra† Jesse R. O’Hanley ‡

Abstract

Optimizing system resilience is concerned with the development of strategies to restore a system to normal operations as quickly and efficiently as possible following potential disruption. To this end, we present in this article a bilevel mixed integer linear program for protecting an uncapacitated median type facility network against worst-case losses, taking into account the role of facility recovery time on system performance and the possibility of multiple disruptions over time. The model differs from previous types of facility protection models in that protection is not necessarily assumed to prevent facility failure altogether, but more precisely to speed up recovery time following a potential disruption. Three different decomposition approaches are devised to optimally solve medium to large problem instances. Computational results provide a cross comparison of the efficiency of each algorithm. Additionally, we present an analysis to estimate cost-efficient levels of investments in protection resources.

Keywords. OR in strategic planning, location, protection, bilevel programming, decomposition.

1 Introduction

In this paper, we consider the problem of reducing the impact of component failures on service and supply systems. Our specific aim is to decide which components to harden or protect, not necessarily with the intent of fully preventing component failures, but more precisely to speed-up the recovery time of the system following a possible worst-case disruption pattern.

∗Kent Business School, University of Kent, CT2 7PE Canterbury, UK. E-mail: cl243@kent.ac.uk
†Corresponding author. Kent Business School, University of Kent, CT2 7PE Canterbury, UK. Tel: +44-1227-824556. Fax: +44-1227-761187. E-mail: m.p.scaparra@kent.ac.uk
‡Kent Business School, University of Kent, CT2 7PE Canterbury, UK. E-mail: j.ohanley@kent.ac.uk
System failure is a highly relevant and timely issue in the design and operation of modern, well-functioning infrastructure networks (Murray and Grubesic, 2007). In practice, component failures may occur for any number of reasons, including equipment breakdowns, industrial accidents (e.g., fires) and even deliberate sabotage or attack (e.g., a terrorist strike). Such failures make a system partially or wholly inoperable for a given length of time and may entail significant direct or indirect costs. This is particularly so in relation to attack and subsequent disruption of critical infrastructure networks (Church et al., 2004; Kamien, 2006) such as electric power grids, transportation hubs and public health facilities.

In many cases, preventive steps can be taken such that when and if failures occur, the ensuing downtime of affected components is reduced. The obvious benefit of this is the improved speed with which the system can be restored to full operational status and, in turn, limit the overall cost that may be incurred during system downtime. At the extreme end, it is sometimes possible to effectively reduce recovery time to zero, in which case a component becomes fully protected from failure and its attendant costs.

General measures designed to avoid disruption and reduce recovery time include adding built-in redundancies, expanding capacity, installation of structural reinforcements and barriers, preventive maintenance, monitoring and inspection (Parry, 1991). In areas prone to flooding, for example, a variety of measures are often taken to prevent failures and or speed-up recovery time: pumps and backup power generators can be put in place; vital road links can be repositioned on elevated terrain; levees and storm drain systems can be built or expanded.

In what follows, we propose a model for protecting an uncapacitated median type system against worst-cases losses, incorporating both facility recovery time and the possibility of multiple disruptions over time. Referred to as the Fortification r-Interdiction Median problem with facility recovery Time (FRIMT), we adopt a defender-attacker or fortification-interdiction type framework (Brown et al., 2006) where a planner seeks to allocate protection resources among facilities in order to reduce the length of time customers must be assigned to more distant facilities than their closest one following a worst-case attack by an interdictor. The interdictor has the ability to strike a fixed number of facilities over a given time horizon. Our model is formulated as a bilevel mixed integer linear program (BMILP), where the upper level models the planner’s protection decisions and the lower level models the interdictor’s optimal response to a given protection strategy.

It should be noted that the use of an interdiction framework does not necessitate the existence of an intelligent attacker. The attacker subproblem is merely used as a device to estimate a worst-case damage scenario for any feasible protection strategy. The model is more widely applicable to problems involving natural disasters when the impacts of disruption are severe.
enough, as is often the case with critical infrastructure systems, to warrant a highly risk-averse decision making criterion based on minimizing the maximum damage.

Recent work on interdiction and fortification based optimization models has tackled a number of issues related to our current problem. In Church et al. (2004) a simple interdiction model is formulated in order to identify the $r$ most critical facilities within an existing $p$-median network whose combined failure would result in the largest increase in total demand-weighted distance. Church and Scaparra (2007) build upon this by introducing fortification into the problem. A BLMIP is developed in which $q$ facilities may be protected in order to minimize the maximum possible damage. Two different solution approaches for this model are discussed in Scaparra and Church (2008a) and Scaparra and Church (2008b). Variations on this basic protection model have subsequently been proposed that include different model facets such as facility capacities (Scaparra and Church, 2010), a random number of possible losses (Liberatore et al., 2010b; Liberatore and Scaparra, 2011), the propagation of disruption over large areas (Liberatore et al., 2010a), capacity expansion and security budget constraints (Aksen et al., 2010), and network flow systems (Cappanera and Scaparra, 2011).

All the above protection models can be casted as multi-level defender-attacker models whose general framework is introduced in Brown et al. (2006). Some practical applications of defender-attacker models to critical infrastructure protection can be found in Brown et al. (2006) for electric power grids, subways, and airports and in Qiao et al. (2007) for water supply networks. Other relevant articles based on game theoretic approaches include Azaiez and Bier (2007), Zhuang and Bier (2007), Jenelius et al. (2010), and Levitin and Hausken (2010).

For the most part, recovery time and the concept of investing protection resources to reduced recovery time have been completely disregarded in the literature, including all of the above. Instead, most models have been constructed as being static, without any thought for the post-recovery phase during which a system is brought back into full operation.

Two exceptions are Losada et al. (2009a) and Holmgren et al. (2007). In Losada et al. (2009a), recovery-time and the availability of multiple attack windows are introduced into the interdiction of a median facility network. Fortification of facilities is not considered. Our present work extends Losada et al. (2009a) by incorporating fortification where different amounts of resources may be invested to reduce facility recovery times to varying degrees. Holmgren et al. (2007) consider the problem of evaluating different protection strategies for electric power grids. Protection resources can be allocated for protection and/or recovery. This model differs from our model in several ways. For example, in Holmgren et al. (2007) all disruptions occur at the same time. Additionally, no solution approach is proposed for
optimizing their bi-level model. Instead protection strategies are evaluated only for a few interdiction scenarios which are defined \textit{a-priori} and only involve one or two components. The protection strategies are then compared with each other according to different criteria.

Besides the model itself with its time-dynamic characteristics, one of our other main contributions is the development of several decomposition based methods for optimally solving FRIMT. Bilevel programs such as FRIMT are notoriously difficult to solve optimally. One widely used approach to solve bilevel models is problem reformulation, which usually involves dual transformation of the lower level (Israeli and Wood, 2002; Lim and Smith, 2007; Losada et al., 2009b) or replacement of the lower level by its KKT optimality conditions (Wang et al., 2000; Arroyo and Galiana, 2005). Given the presence of integer variables in the lower level of FRIMT, however, such approaches cannot be readily applied.

The literature on bilevel problems with integer variables in the lower level suggests that general solution approaches are inefficient at solving instances with a realistic number of variables (e.g., Moore and Bard (1990)). Other general approaches presented in Caroe and Tind (1998) and Sherali and Fraticelli (2002) have not been tested on a wide variety of problems, making it unclear as to their practical efficiency. Gabriel et al. (2010) suggest a general method based on Benders decomposition and solve instances of moderate size. However, their solution approach does not guarantee the optimality of a solution. Other general and problem specific approaches for solving bilevel models can be found in Labbé et al. (1998); Scaparra and Church (2008a); Garg and Smith (2008); Taçen et al. (2009) and O’Hanley and Church (2011).

To solve FRIMT both optimally and efficiently, we propose three different exact decomposition algorithms. In all three, FRIMT is split into two interlinked subproblems: an upper restricted master problem (RMP) and a lower level interdiction subproblem (SP). Each protection strategy found by the RMP is fed into the SP to determine an optimal interdiction pattern. The result is a feasible bilevel solution to FRIMT. Cuts are then generated based on solution to the SP and then added to the RMP. The methods each differ with respect to the specific form of the RMP and the type of cuts that are generated.

In the first decomposition method, a Benders type approach is employed whereby \textit{Adaptive Benders optimally Cut} (ABCs) are used to progressively tighten an upper bound derived from the RMP. Each ABC computes an exact value for an interdiction pattern found by the SP in previous iterations. The cuts are adaptive in that the dual variables used to define a Benders cut are not fixed but are directly optimized over during solution of the RMP. Unlike previous approaches (Gabriel et al., 2010), ours is guaranteed to converge to a proven optimal solution. In the second approach, the RMP produces no bounds and is simply a collection of
feasibility cuts, known as *Supervalid Inequalities* (SVIs) (Israeli and Wood, 2002; O’Hanley and Church, 2011). SVIs, unlike regular valid inequalities, remove integer solutions from the feasible space but are guaranteed not to remove all optimal solutions unless one has already been found. The SVIs are used here as way of forcing the RMP away from clearly dominated solutions. The algorithm terminates when a sufficient number of SVIs have been added to make the RMP infeasible. Lastly, we propose a hybrid decomposition method which relies on the combined use of ABCs and SVIs.

The remainder of the paper is organized as follows. In Section 2, we revisit the interdiction problem with recovery times first presented in Losada et al. (2009a) and propose a much more efficient and scalable formulation. In Section 3, we present the bilevel formulation of our model FRIMT. Section 4 provides details of the three decomposition methods proposed to solve FRIMT. In Section 5, we compare the computational performance of the three methods on several problem instances. We also examine how changes in the protection budget impact overall system efficiency with the aim of identifying cost-effective levels of investment in protection. Finally, in Section 6 we provide a summary of the main contributions of the paper, as well as discuss some problem insights and suggested areas of future research.

2 A reformulation for the \( r \)-Interdiction Median problem with facility recovery times and frequent disruptions

The *r*-Interdiction Median problem with facility recovery Times and frequent disruptions (RIMT) was first presented in Losada et al. (2009a). The basic modeling assumptions of RIMT are as follows. A supply system with \( p \) uncapacitated facilities provides service to \( n \) customers. Each customer is served by his closest operating facility unless the closest facility is out of service due to some kind of failure or disruption. Each facility has a recovery time associated with it, which denotes the time for the facility to return to normal operation after a failure. A disrupted facility is completely inoperable throughout its entire recovery time so that no customers can be assigned to it. Disruptions can occur at any time period within the planning horizon and the same facility can be disrupted multiple times. At most \( r \) disruptions may take place during the planning horizon. The model identifies the worst-case disruption scenario by maximizing the total demand weighted travel distance in serving the customers over all time periods. In this sense, RIMT explicitly optimizes both the level and duration of disruption.

The formulation we propose for RIMT differs from the one given in Losada et al. (2009a) and is based on the following observation. If the customer demands are homogeneous throughout
the planning horizon, the exact time period when the interdiction of each facility takes place is irrelevant. A worst-case disruption scenario can always be identified by simply determining how many times each facility is interdicted, regardless of the time of interdiction.

This observation allows us to reformulate RIMT in a more compact and efficient way. A formal discussion of the equivalence of the two formulations is presented at the end of this section.

Let $N$ be the set of customers indexed by $i$ and $F$ the set of facilities indexed by $j$ or $i_k$, where $i_k$ denotes the $k$-th closest facility to customer $i$. Also, let $d_{ij}$ be the distance between customer $i$ and facility $j$ and $h_i$ be the demand of customer $i$. In addition, let $r$ be the interdiction budget, $T$ be the number of time periods in the planning horizon and $G_j$ be the recovery time of facility $j$. Finally, let $s_j \geq 0$ and $x_{ij} \geq 0$ be the interdiction and the assignment variables, respectively. The variable $s_j$ denotes the number of times facility $j$ is interdicted during the planning horizon. The variable $x_{ij}$ represents the total number of times customer $i$ is served by facility $j$ during all $T$ time periods.

The revisited RIMT model (RIMT-2) is:

\begin{align}
\text{max} & \quad \sum_{i \in N} \sum_{j \in F} h_i d_{ij} x_{ij} \\
\text{s.t.} & \quad \sum_{j \in F} x_{ij} = T, \forall i \in N \\
& \quad \sum_{j \in F} s_j \leq r \\
& \quad \sum_{l=k+1}^{\vert F \vert} x_{ii_l} \leq G_{i_k} s_{i_k}, \forall i \in N, 1 \leq k \leq \vert F \vert - 1 \\
& \quad s_j \in \mathbb{Z}_+, \forall j \in F \\
& \quad x_{ij} \in \mathbb{Z}_+, \forall i \in N, j \in F
\end{align}

The objective function maximizes the sum of the weighted distances of serving all customers over all time periods. Constraints (2) state that each customer must receive service in each time period. Constraint (3) limits the total number of disruptions of the system to be no more than $r$. Constraints (4) indicate that the maximum number of times that customer $i$ is served by facilities further than $i_k$ is $G_{i_k} s_{i_k}$ times. If facility $i_k$ is not interdicted (i.e., $s_{i_k} = 0$), then the allocation of service to facilities further than the $k$th closest facility is not permitted. Constraints (5) and (6) state that the decision variables are integer. Finally, to tighten our formulation we add some upper bounds on the $s_j$ variables: $s_j \leq \lceil T/G_j \rceil, \forall j \in F$.  

6
These constraints state that there is no gain from interdicting facility $j$ more times than the number of times required to put $j$ out of service over the entire planning horizon (i.e., for $T$ time periods). Note that because of the problem structure, the assignment variables $x_{ij}$ automatically takes on integer values. Thus, for computational purposes $x_{ij}$ can be redefined as nonnegative continuous variables: $x_{ij} \geq 0, \forall i \in N, j \in F$.

We can also reduce the number of the assignment variables $x_{ij}$ using a preprocessing step. The idea is based on the fact that any customer will travel in the worst-case to his $r+1$ closest facility, $i_{r+1}$. Hence, we set to zero all assignment variables with facility $j$ at position $r+2$ or further from customer $i$: $x_{ijk} = 0, \forall i \in N, k > r + 1$.

The equivalence between the original RIMT formulation and our new model (1)-(6) is based on the following assumption and proposition.

**Assumption 1.** (a) All the facilities are hit for the first time at $t = 0$. (b) If a facility is hit more than once, disruption of the facility occurs consecutively, with one starting as soon as the impact of the former disruption ends.

**Proposition 1:** Let the $H^1$ and $H^2$ be the objective function values for [RIMT] associated with any two feasible solutions $y^1 = (s^1, x^1)$ and $y^2 = (s^2, x^2)$, where the set of disrupted facilities as well as the number of times each facility is disrupted are the same for both $y^1$ and $y^2$, but differ possibly with respect to when disruptions occur. If $y^2$ satisfies Assumption 1, then $H^2 \geq H^1$.

**Proof:** Given a solution $y$ to RIMT, let $c_j(y)$ be the number of periods throughout the planning horizon that facility $j$ is out of service. Further, let $p_{ik}(y)$ represent the number of times, given any solution $y$, that customer $i$ is served by a facility further away than facility $i_k$. An upper bound on $p_{ik}(y)$, denoted $\bar{p}_{ik}(y)$, is given by $\bar{p}_{ik}(y) = \min_{l \leq k} \{c_i(y)\}$. To prove that $H^2 \geq H^1$, it is sufficient to show that $p_{ik}(y^2) \geq p_{ik}(y^1), \forall i \in N, 1 \leq k \leq |F|$.

By assumption, $c_j(y^1) = c_j(y^2)$ for each facility $j$, hence $\bar{p}_{ik}(y^1) = \bar{p}_{ik}(y^2)$. Furthermore, since $y^2$ satisfies Assumption 1, it follows that $p_{ik}(y^2) = \bar{p}_{ik}(y^2), \forall i \in N$, given that facility $i_k$ and all those closer than it (i.e., $i_l$ for all $l \leq k$) are collectively disrupted a total of $\min_{l \leq k} \{c_i(y)\}$ periods. Finally, based on the fact that $\bar{p}_{ik}(y) \geq p_{ik}(y)$, we find: $p_{ik}(y^2) = \bar{p}_{ik}(y^2) = \bar{p}_{ik}(y^1) \geq p_{ik}(y^1)$ $\square$

Note that by disregarding the time of interdiction and only considering the frequency (and hence the duration) of disruption, formulation RIMT-2 implicitly satisfies Assumption 1. Consequently, its feasible region is significantly smaller than the one of the original formulation but still contains at least one optimal solution to RIMT, according to Proposition 1. A detailed computational comparison between the original RIMT and RIMT-2, as well as a
practical example of Proposition 1, can be found in Losada (2010). Computational results show that our new reformulation is up to three orders of magnitude faster than that of the initial formulation. This is a considerable gain considering that RIMT is used within the FRIMT bi-level structure to estimate worst-case losses in response to a set of protection strategies. To find an optimal solution to FRIMT using any of our decomposition methods, the RIMT problem needs to be solved repeatedly. Thus, any improvement in the computational time for solving RIMT is amplified when solving the protection problem FRIMT.

3 Bilevel formulation of the fortification problem

Consider the following additional notation. Let $q$ be the protection budget and $m_j$ stand for the reduction in recovery time $G_j$ for every unit of $q$ invested in facility $j$. The decision variables used are as follows. Let $z_j \in \mathbb{Z}_+$ be the amount of protection resources invested in facility $j$, $\tau_j \in \mathbb{Z}_+$ be the total reduction in recovery time of facility $j$, $s_j \in \mathbb{Z}_+$ be the number of times that facility $j$ is disrupted and $x_{ij} \in \mathbb{Z}_+$ be the number of times that customer $i$ is served by facility $j$. A bi-level formulation of FRIMT is:

\[
\begin{align*}
\text{min}_{z, \tau} & \quad H(z, \tau, s, x) \\
\text{s.t.} & \quad \sum_{j \in F} z_j \leq q \quad (8) \\
& \quad z_j \leq \lceil G_j/m_j \rceil, \forall j \in F \quad (9) \\
& \quad \tau_j \leq G_j, \forall j \in F \quad (10) \\
& \quad \tau_j \leq m_j z_j, \forall j \in F \quad (11) \\
& \quad z_j \in \mathbb{Z}_+, \forall j \in F \quad (12) \\
& \quad \tau_j \in \mathbb{Z}_+, \forall j \in F \quad (13) \\
H(z, \tau, s, x) &= \max_{s, x} \sum_{i \in N} \sum_{j \in F} h_i d_{ij} x_{ij} \quad (14) \\
\text{s.t.} & \quad \sum_{j \in F} x_{ij} = T, \forall i \in N \quad (15) \\
& \quad \sum_{j \in F} s_j \leq r \quad (16) \\
& \quad \sum_{i=1}^{|F|} x_{ii_k} \leq (G_{ik} - \tau_{ik}) s_k, \forall i \in N, 1 \leq k \leq |F| - 1 \quad (17)
\end{align*}
\]
\[
\begin{align*}
  s_j & \in \mathbb{Z}_+, \forall j \in F \\
  x_{ij} & \in \mathbb{Z}_+, \forall i \in N, j \in F
\end{align*}
\] (18) (19)

The upper level is given by (7) – (13), while the the lower level is given by (14)-(19). Constraint (8) is a cardinality constraint on the upper level decision variables, which limits the amount of protection resources invested in all facilities to be less than or equal to the protection budget \(q\). Constraints (9) and (10) establish an upper bound on the amount of protection resources spent on each facility and on the maximum reduction in recovery time of each facility, respectively. Constraints (11) link the variables \(z_j\) and \(\tau_j\). Constraints (10) and (11), in conjunction with variables \(\tau_j\), serve to prevent having negative values on the right-hand sides of constraints (17) when \([G_j/m_j]m_j > G_j\), (i.e., the reduction in recovery time is at most the full recovery time). In the lower level, constraints (15) state that each customer must be served by one facility in each time period. Constraints (16) are cardinality constraints on the lower level decision variables, which limit the number of facility disruptions to be no more than \(r\). Constraints (17) indicate that each time facility \(i_k\) is disrupted, customer \(i\) must be served by a facility further than \(i_k\) at most \((G_i - \tau_i)\) times. If facility \(i_k\) is not interdicted, the allocation of service to facilities further than \(i_k\) is not permitted. Constraints (12), (13), (18) and (19) impose integer restrictions on the decision variables. In practice, only constraints (12) and (18) are imposed, given that if \(z\) and \(s\) are integer, the \(\tau\) and \(x\) variables automatically take on positive integer values. Finally, we include the redundant constraints (20) in the lower level to yield a stronger formulation.

\[
s_j \leq \lceil T/(G_j - \tau_j) \rceil, \forall j \in F
\] (20)

We give below a formal definition of the different regions and sets of FRIMT according to Saharidis and Ierapetritou (2009), to which we will refer throughout the following sections. We let \(e\) be a column vector of ones of suitable size and \(I\) be an identity matrix of suitable size. The term \((a \cdot b)\) denotes the inner product between vectors \(a\) and \(b\). We define vectors \(B\) and \(a\) as \(B_j = [G_j/m_j]\) and \(a_j = m_j z_j, \forall j \in F\), respectively.

- **Constraint region:** \(\Omega = \{(z, \tau, s, x) : z \in Z, \tau \in \Gamma, s \in S, x \in X, h(\tau, s) \leq 0, g(\tau, s, x) \leq 0\}\), where \(Z = \{z \in \mathbb{Z}^{\lvert F \rvert} \mid e \cdot z \leq q, z \leq B\}, \Gamma = \{z \in \mathbb{Z}^{\lvert F \rvert} \mid z \leq G, \tau \leq a, G^T = (G_1, G_2, ..., G_{\lvert F \rvert})\}, S = \{s \in \mathbb{Z}^{\lvert F \rvert} \mid e \cdot s \leq r\}, X = \{x \in \mathbb{R}^{\lvert N \rvert \times \lvert F \rvert} \mid xe = \text{diag}(TI)\}\) and where \(g(\tau, s, x) \leq 0\) represents constraints (17) and \(h(\tau, s) \leq 0\) is defined as constraints (20).

- **The projection of \(\Omega\) onto the leader’s decision space:** \(\Omega(Z, \Gamma) = \{z \in Z, \tau \in \Gamma :\)

We include the redundant constraints (20) in the lower level to yield a stronger formulation.
\[ \exists (s, x), (z, \tau, s, x) \in \Omega \] 

- The follower’s feasible region for a given protection strategy \((\hat{z}, \hat{\tau}) \in \Omega(Z, \Gamma) : \Omega(\hat{z}, \hat{\tau}) = \{(s, x) : s \in S, x \in X, h(\tau, s) \leq 0, g(\tau, s, x) \leq 0\} \]

- The follower’s rational set: \(M(\hat{z}, \hat{\tau}) = \{(s, x) \in \Omega(\hat{z}, \hat{\tau}) : \arg \max (H(\hat{z}, \hat{\tau}, s, x))\} \)

- The inducible region (IR): \[ \{(z, \tau, s, x) : (z, \tau) \in \Omega(Z, \Gamma), (s, x) \in M(z, \tau)\} \]

The follower’s feasible set \(\Omega(\hat{z}, \hat{\tau})\) consists of the values of the interdiction variables \(s\) and allocation variables \(x\) that meet the constraints in the lower level program for a given protection strategy \((\hat{z}, \hat{\tau}) \in \Omega(Z, \Gamma)\). An important property of this problem is that for any solution from the leader’s decision space, there exists a feasible solution in the follower’s feasible set (i.e., \(M(\hat{z}, \hat{\tau}) \neq \emptyset, \forall (\hat{z}, \hat{\tau}) \in \Omega(Z, \Gamma)\)). The follower’s rational set \(M(\hat{z}, \hat{\tau})\) contains the values of variables \(s\) and \(x\) that maximize the objective function of the lower level program for a given protection strategy \((\hat{z}, \hat{\tau}) \in \Omega(Z, \Gamma)\).

In what follows, we refer to the leader’s decisions or to protection strategies as the solutions contained in \(\Omega(Z, \Gamma)\). Likewise, we refer to the follower’s response or to interdiction strategies for a given protection strategy as the values of the interdiction variables \(s\) such that \((s, x) \in M(\hat{z}, \hat{\tau}), \hat{z} \in Z, \hat{\tau} \in \Gamma\).

## 4 Decomposition methods

This section provides a detailed description of the proposed exact decomposition methods for FRIMT: Benders decomposition (D-Bend), SVI based decomposition (D-SVI) and a hybrid decomposition (D-H). While the subproblem (SP) is the same for all decomposition methods, the specific form of the restricted master problem (RMP), the type of cutting planes and the convergence criteria all differ.

By solving the RMP, a feasible protection strategy \((\hat{z}, \hat{\tau}) \in \Omega(Z, \Gamma)\) is found which is used to update the values of the facility recovery times in the next SP solved, denoted for simplicity as SP(\(\hat{\tau}\)). SP(\(\hat{\tau}\)) is an equivalent version of RIMT-2 resulting from using updated values of \(G\). By solving SP(\(\hat{\tau}\)) we obtain two useful pieces of information: (1) the most harmful disruption strategy if SP(\(\hat{\tau}\)) is solved to optimality (the follower’s response) and (2) a feasible solution to FRIMT, where the former can be used to generate a cutting plane for the RMP, while the latter provides an upper bound to FRIMT.

The outline of this section is as follows. In Section 4.1 we define the SP for a given protection strategy. Sections 4.2, 4.3 and 4.4 describe D-Bend, D-SVI, and D-H, respectively. Each of
these sections provides a description of the specific RMP, the type of cutting planes used, an enumeration of the steps in each algorithm and a general discussion of each algorithm.

4.1 Lower level subproblem (SP)

Let \( G'_{j}(\hat{\tau}) = G_{j} - \hat{\tau}_{j} \) be the updated recovery time of facility \( j \) for a given protection strategy, \((\hat{z}, \hat{\tau}) \in \Omega(Z, \Gamma)\). Then SP(\(\hat{\tau}\)) is given by the following,

\[
[\text{SP}(\hat{\tau})] \quad H(\hat{z}, \hat{\tau}, \hat{s}, \hat{x}) = \max_{s, x} \sum_{i \in N} \sum_{j \in F} h_{ij}d_{ij}x_{ij}
\]

s.t. \[
\sum_{j \in F} x_{ij} = T, \forall i \in N
\]

\[
\sum_{j \in F} s_{j} \leq r
\]

\[
|F| \sum_{l=k+1}^{l} x_{ii_{l}} \leq G'_{i_{k}}(\hat{\tau})s_{i_{k}}, \forall i \in N, 1 \leq k \leq |F| - 1
\]

\[
s_{j} \leq \lceil T/G'_{j}(\hat{\tau}) \rceil, \forall j \in F
\]

along with constraints (18) and (19). SP(\(\hat{\tau}\)) is conceptually equivalent to RIMT-2 but with updated values of the facility recovery times \( G'_{j}(\hat{\tau}) \). The optimal solution to \( \text{SP}(\hat{\tau}) \), \((\hat{s}, \hat{x})\), is also a feasible solution to FRIMT (i.e., \((\hat{s}, \hat{x}) \in M(\hat{z}, \hat{\tau})\)). The objective function value associated with the feasible solution \((\hat{z}, \hat{\tau}, \hat{s}, \hat{x}) \in IR, H(\hat{z}, \hat{\tau}, \hat{s}, \hat{x})\) provides an upper bound to the objective function value of the optimal solution to FRIMT.

4.2 Benders decomposition (D-Bend)

A classical application of Benders decomposition (Benders, 1962; Geoffrion, 1972) to FRIMT is not possible as the lower level includes integer variables, making the dual variables of the lower level not readily available. Some authors have adapted Benders decomposition to overcome the problem of having a non-convex lower level created by the presence of discrete variables (Gabriel et al., 2010; Caroe and Tind, 1998; Sherali and Fraticelli, 2002). Gabriel et al. (2010), for example, propose an algorithm that consists of partitioning the domain of the complicating integer-valued variables (decision variables for the RMP) in such a way that within these regions the resulting problems optimize over convex regions. They then apply classical Benders decomposition to solve each problem associated with a given domain of the complicating variables to obtain a local minimum (for a minimization problem). Since the
global minimum of the problem must lie in a convex subregion, one can select the smallest identified local minimum which must also be a global minimum (Delgadillo et al., 2010). The optimality of the procedure is predicated on the assumption that the partitioning of the domain does, in fact, lead to problems that optimize over convex regions. Unfortunately, they use necessary but not sufficient conditions for partitioning the complicating variables into convex regions. Therefore, their solution approach fails to guarantee the optimality of the solutions obtained. In practice, however, it finds the optimal solution in most reported tests of moderately sized problems.

In D-Bend, we seek to design an RMP that, at a given iteration $w$, identifies a protection strategy $(\hat{z}, \hat{\tau})^w$, which minimizes the maximum impact generated by those disruptions identified by the SPs solved throughout the first $w$ iterations. In the context of FRIMT, specialized Benders optimality cuts are used to evaluate the maximum impact generated by any disruption pattern identified by the SP throughout the first $w$ iterations. Unlike classical Benders cuts, our approach has the additional advantage that the cutting planes generated do not eliminate the global minimum. Specifically, at a given iteration, we solve $SP(\hat{\tau})$ for an optimal interdiction pattern $\hat{s}$ and then fix the integer variables of $SP(\hat{\tau})$ to their optimal values. Up to this point we follow the methodology of Gabriel et al. (2010).

Now, instead of solving $SP(\hat{\tau}, \hat{s})$ for the optimal value of its dual variables, we take the dual of $SP(\hat{\tau}, \hat{s})$ and, without solving it, we introduce it in toto directly into the RMP. In this way, the RMP optimizes over its initial set of complicating variables plus the dual variables of $SP(\hat{\tau}, \hat{s})$. We call the dual of $SP(\hat{\tau}, \hat{s})$ an adaptive Benders optimality cut (ABC).

By not fixing the dual variables of $SP(\hat{\tau}, \hat{s})$, the RMP is able to generate a cut that exactly represents the value of the overall demand weighted travel distance when disruption pattern $\hat{s}$ and protection strategy $(\hat{z}, \hat{\tau})$ are implemented. In other words, the assignment of customers to available facilities is optimal for a given disruption and protection pattern. This is not the case for the Benders cuts proposed by Gabriel et al. (2010). If we generated a Benders-like cut like the one proposed by Gabriel et al. (2010) for FRIMT, we would have, for fixed $\hat{s}$ and $(\hat{z}, \hat{\tau})$, the overall demand weighted travel distance computed by the active cutting planes representing an upper bound on the real overall travel distance. This is the reason why the ABCs allow D-Bend to converge optimally, unlike the cuts of Gabriel et al. (2010).

**MRP and cutting planes in D-Bend**

Formally, the procedure of generating ABCs is implemented as follows. At a given iteration, we solve $SP(\hat{\tau})$ and obtain $\hat{s}$. Note that $\Omega(z, \tau) \neq \emptyset, \forall (z, \tau) \in \Omega(Z, \Gamma)$ (i.e., for every protection strategy there exists a feasible solution of the follower’s response). Thus, there is
no need to add Benders feasibility cuts to the RMP. Then, given \((\hat{z}, \hat{\tau})\) and \(\hat{s}\) we consider \(SP(\hat{\tau}, \hat{s})\), defined as:

\[
\text{max}_x \quad \sum_{i \in N} \sum_{j \in F} h_{ij} d_{ij} x_{ij} \quad (26)
\]

s.t. \[
\sum_{j \in F} x_{ij} = T, \forall i \in N \quad (27)
\]

\[
\sum_{l=k+1}^{\mid F \mid} x_{ii_l} \leq (G_{ik} - \hat{\tau}_{ik}) \hat{s}_{ik}, \forall i \in N, 1 \leq k \leq |F| - 1 \quad (28)
\]

\[
x_{ij} \geq 0, \forall i \in N, j \in F \quad (29)
\]

\(SP(\hat{\tau}, \hat{s})\) is used to optimally assign customers to facilities given predefined protection and interdiction patterns. The feasible domain of \(SP(\hat{\tau}, \hat{s})\) is convex and continuous for given values of the protection and interdiction strategies. Therefore, we can take the dual of \(SP(\hat{\tau}, \hat{s})\). Let \(\beta_i\) and \(\mu_{ii_k}\) be the dual variables associated with constraints (27) and (28), respectively. The dual of the maximization subproblem (26)-(29) is:

\[
\text{min}_{\mu, \beta} \quad \sum_{i \in N} \left( \sum_{k=1}^{\mid F \mid - 1} \mu_{ii_k} (G_{ik} - \hat{\tau}_{ik}) \hat{s}_{ik} + T \beta_i \right) \quad (30)
\]

s.t. \[
\sum_{l=1}^{k-1} \mu_{ii_l} + \beta_i \geq h_{ij} d_{ii_k}, \forall i \in N, 1 < k \leq |F| \quad (31)
\]

\[
\beta_i \geq h_{ij} d_{ii_1}, \forall i \in N \quad (32)
\]

\[
\mu_{ii_k} \geq 0, \forall i \in N, 1 \leq k \leq |F| - 1 \quad (33)
\]

The program DSP(\(\hat{s}\)) is what we call an adaptive Benders cut (ABC). At each iteration, an ABC is introduced into the RMP. Note that, as mentioned above, for the purpose of taking the dual of the SP, we consider the decision variables \(\tau\) as parameters to DSP(\(\hat{s}\)). However, when we add DSP(\(\hat{s}\)) to the RMP we obtain a more general problem over variables \(z, \tau, \mu, \beta\) (only the \(s\) are fixed). Hence, we have to resolve the nonlinearity in the objective function (i.e., we have to linearize the \(\mu_{ii_k} \tau_{ik}\) terms). The linearization of these terms is not trivial, given that the variables \(\tau_{ik}\) can take on any integer value in the range \([0, G_{ik}]\). The linearization we employ is described in Appendix A. The resulting formulation of the RMP at iteration \(w\), denoted as RMP-D-Bend, is given in Appendix B.

D-Bend converges to optimality when the lower bound (i.e., the objective function of the RMP) is greater than or equal to the upper bound. Since the number of feasible solutions to
the SP for any feasible protection strategy is finite, convergence must be attained in a finite number of iterations.

At iteration $w$, let $\hat{W}_w = \{ s^b | \exists s^b, (s^b, x^b) \in M(\hat{z}, \hat{\tau}), (\hat{z}, \hat{\tau}) \in \Omega(Z, \Gamma), 1 \leq b \leq w \}$ be the set of the follower’s responses (interdiction patterns) found in the first $w$ iterations of the algorithm. In addition, we consider the following notation.

**Notation:**

- $H^{up} =$ Best upper bound of FRIMT
- $H^{low} =$ Lower bound of FRIMT
- $H(\hat{\tau}^{w-1}) =$ Objective function value associated with the solution of SP($\hat{\tau}^{w-1}$), given protection strategy $(\hat{z}, \hat{\tau})^{w-1} \in \Omega(Z, \Gamma)$ identified by the RMP in iteration $w - 1$.
  (For simplicity we use $H(\hat{\tau}^{w-1})$ instead of $H(\hat{z}^{w-1}, \hat{\tau}^{w-1}, s^w, x^w)$)
- $H_{Heur} =$ Objective function value associated with a heuristically generated solution of the RMP
- $\epsilon =$ Allowed optimality gap for an instance of FRIMT
- $(z, \tau, s, x)^+ \in IR =$ Solution associated with the best upper bound of FRIMT ($H^{up}$)
- $(z, \tau, s, x)^* \in IR =$ Solution associated with an optimal solution of FRIMT

**D-Bend Algorithm**

**Step 0.** Initialize: $H^{up} = \infty, H^{low} = -\infty, \hat{W}_0 = \emptyset, (\hat{z}, \hat{\tau})^0 \leftarrow 0, (z, \tau, s, x)^* \leftarrow 0, w = 1$

**Step 1.** Solve [SP($\hat{\tau}^{w-1}$)] for a follower’s response $(s^w, x^w) \in M((\hat{z}, \hat{\tau})^{w-1})$ and associated $H(\hat{\tau}^{w-1})$:

(i) $\hat{W}_w \leftarrow \hat{W}_{w-1} \cup s^w$

(ii) If $H(\hat{\tau}^{w-1}) < H^{up}$ then $H^{up} \leftarrow H(\hat{\tau}^{w-1})$ and $(z, \tau, s, x)^+ \leftarrow (z^{w-1}, \tau^{w-1}, s^w, x^w)$

(iii) $w = w + 1$

**Step 2.** Given $\hat{W}_w$:

(i) Solve RMP-D-Bend heuristically for $(\hat{z}, \hat{\tau})^w \in \Omega(Z, \Gamma)$ and $H_{Heur}$. If $H_{Heur} < H^{up}$, go to Step 1. Else, go to (ii).

(ii) Solve RMP-D-Bend exactly for $(\hat{z}, \hat{\tau})^w$ and $H^{low}$

**Step 3.** If $H^{up} - H^{low} < \epsilon$, an optimal solution has been found with optimality gap $\epsilon$, go to Step 4. Else, go to Step 1.

**Step 4.** Print optimal solution of FRIMT: $(z, \tau, s, x)^* = (z, \tau, s, x)^+$
Note that in Step 2, we first try to find a feasible protection strategy by solving RMP-D-Bend heuristically and use the obtained solution \((\hat{z}, \hat{\tau})\in \Omega(Z, \Gamma)\) for the subsequent [SP(\hat{\tau})], provided that \(H_{Heur} < H^{up}\) (Israeli and Wood, 2002). That is, if \(Z_{Heur} \leq H^{up}\) we use the solution \((\hat{z}, \hat{\tau})\in \Omega(Z, \Gamma)\) to generate an ABC but we do not update the best lower bound. We only update \(H^{low}\) in Step 2 (ii) when [RMP-D-Bend] is solved to optimality. The heuristic used in Step 2(i) is a simple greedy heuristic. Preliminary results demonstrated that the use of this heuristic within the D-Bend framework helped to reduce the computing time of the algorithm. To further improve the computational efficiency of D-Bend, we also devised some variable bounding and reduction mechanisms. These and the related proofs of correctness are provided in Appendix C.

4.3 SVI based decomposition (D-SVI)

The RMP in D-SVI is essentially a feasibility seeking problem consisting of a set of cutting planes called Supervalid Inequalities (SVIs). These are not valid inequalities in a strict sense, as they necessarily remove feasible integer solutions. Nonetheless, they reduce the size of the feasible region of the RMP by making the incumbent solution and possibly other solutions infeasible in subsequent iterations. Importantly, however, an SVI is guaranteed not to eliminate all optimal solutions unless an optimal solution has already been found (O’Hanley and Church, 2011).

The Super Valid Inequalities (SVIs)

The rational behind the SVI based scheme is that if the leader wants to decrease the objective function value of the follower, he/she is bound to protect facilities in such a way that previously “seen” interdiction strategies must become less attractive in the future. Specifically, the leader either needs to protect at least one of the disrupted facilities where no protection resources were previously allocated or needs to increase the amount of protection resources of at least one of the disrupted facilities already being protected.

Formally, for a given response of the follower at iteration \(w\), \((s^w, x^w) \in M((\hat{z}, \hat{\tau})^{w-1})\), let the set \(SZ = \{j \mid s^w_j \geq 1, \hat{z}^{w-1}_j \geq 1, \forall j \in F\}\) contain the indices of the disrupted facilities which are protected in the protection strategy \((\hat{z}, \hat{\tau})^{w-1}\). Also, let set \(S^+ = \{j \in F : s^w_j \geq 1, \hat{z}^{w-1}_j = 0\}\) contain the indices of the facilities that are disrupted but not protected.

After solving the SP at iteration \(w\), we distinguish three distinct cases for which a set of SVIs, denoted \(SVI_w\), can be generated:
**Case 1.** None of the disrupted facilities are protected: $SZ = \emptyset$. In this case $SV I_w$ contains a single inequality, $\sum_{j \in S^+} z_j \geq 1$, which states that at least one disrupted facility needs to be protected.

**Case 2.** Some disrupted facilities are protected: $SZ \neq \emptyset$ and $S^+ \neq \emptyset$. Here, a set of either-or inequalities is generated. Specifically, to improve on the follower’s subsequent response, either at least one facility in $S^+$ must be protected or the protection resources for at least one facility in $SZ$ must be increased. Formally, either $\sum_{j \in S^+} z_j \geq 1$ must be true or $z_j \geq \hat{z}_j^{w-1} + 1$ must be true, for some $j \in SZ$. Let $\phi \in \{0,1\}^{1+|SZ|}$ be a vector of auxiliary variables, where each auxiliary variable is associated with one of the either-or constraints. An inequality is active if $\phi_l = 1$, $1 \leq l \leq |SZ|+1$, and only one inequality can be active, $\sum_{l=1}^{|SZ|+1} \phi_l = 1$. Also, let $l_j$ be the position in the vector of variables $\phi$ corresponding to facility $j \in SZ$, where $2 \leq l_j \leq |SZ| + 1$. Then, $SV I_w$ at iteration $w$, is:

\[
\sum_{j \in S^+} z_j - 1 \geq -M_1(1 - \phi_1)
\]

\[
z_j - \hat{z}_j^{w-1} - 1 \geq -M_l_j (1 - \phi_{l_j}), \forall j \in SZ
\]

\[
|SZ|+1 \sum_{l=1} \phi_l = 1
\]

where $M_1$ and $M_l_j$ are sufficiently large numbers. Possible values are: $M_1 = M_l_j = q$.

**Case 3.** All the disrupted facilities are protected: $SZ \neq \emptyset$ and $S^+ = \emptyset$. Here, we also have a set of either-or inequalities. These consist of inequalities (35) and (36) with $\phi_1 = 0$.

The proposition below shows that the set of inequalities described above satisfy the condition of being super-valid (O’Hanley and Church, 2011).

**Proposition 2:** For any and all Cases 1-3, the set of cutting planes $SV I_w$ is supervalid.

**Proof:** If the incumbent bilevel solution is optimal, $(\hat{z}, \hat{\tau}, \hat{s}, \hat{x})^*$, then by definition the set of constraints $SV I_w$ is supervalid. Given a sub-optimal solution found at iteration $w$, $(\tilde{z}, \tilde{\tau}, \tilde{s}, \tilde{x})^w = (\hat{z}, \hat{\tau}, \hat{s}, \hat{x})^+\hat{z}$, the addition of $SV I_w$ generates a new protection strategy $(\hat{z}, \hat{\tau})^w \neq (\tilde{z}, \tilde{\tau})^{w-1}$ that will either prevent the follower’s former response $s^w$ or if $s^w$ is found again the objective function value will improve because additional protection resources where...
invested in one or more disrupted facilities, i.e., either $s_{w+1} \neq s_w$ or if $s_{w+1} = s_w$, then $H(\hat{\tau}^w) < H(\hat{\tau}^{w-1})$. Thus, $SVI_w$ is guaranteed to eliminate the incumbent solution to the RMP for any iteration $w$, i.e., $(z, \tau, s, x)^w \neq (z, \tau, s, x)^{w-1}$. □

The RMP for D-SVI
The RMP in D-SVI is a feasibility seeking problem. That is, it does not have an objective function and only consists of a set of SVIs, which are iteratively added throughout the execution of the algorithm. Therefore, upon solving the RMP, one stops as soon as a feasible solution is found. This is an important property of D-SVI which makes the solution of the RMP very fast and, overall, greatly speeds up convergence, compared to D-Bend.

D-SVI Algorithm
The mechanics of the algorithm of D-SVI are similar to that of D-Bend. We solve iteratively the SP and the RMP, where the solution of the SP is used to generate cutting planes for the RMP. That is, at each iteration $w$ of the algorithm D-SVI, the solution of $\text{SP}(\hat{\tau}^w_{w-1})$ is used to generate a set of super valid inequalities $SVI_w$. Let the set $\hat{SVI}_w$ be a cumulative set of SVIs (i.e., $\hat{SVI}_w = \hat{SVI}_{w-1} \cup SVI_w$). Then, the RMP at iteration $w$ is a satisfiability problem defined by: $Z \cup \Gamma \cup \hat{SVI}_w$ (see Section 3 for the definition of constraint sets $Z$ and $\Gamma$).

The major difference between D-SVI and D-Bend is the convergence criterion. As we do not have an objective function value for the RMP in D-SVI, a solution can be guaranteed to be optimal for FRIMT only when the RMP becomes infeasible, meaning no protection strategy can thwart all of the interdiction patterns identified so far.

The D-SVI algorithm is:

Step 0. Initialize: $H_{\text{up}} = \infty$, $H_{\text{low}} = -\infty$, $\hat{SVI}_0 = \emptyset$, $z^0 \leftarrow 0$, $(\hat{z}, \hat{\tau})^0 \leftarrow 0$, $(z, \tau, s, x)^* \leftarrow 0$, $w = 1$

Step 1. Solve $\text{SP}(\hat{\tau}^{w-1})$ for a follower’s response $(s^w, x^w) \in M((\hat{z}, \hat{\tau})^{w-1})$ and associated $H(\hat{\tau}^{w-1})$:

(i) Generate $SVI_w$ and update $\hat{SVI}_w \leftarrow \hat{SVI}_{w-1} \cup SVI_w$

(ii) If $H(\hat{\tau}^{w-1}) < H_{\text{up}}$ then $H_{\text{up}} \leftarrow H(\hat{\tau}^w)$ and $(z, \tau, s, x)^+ \leftarrow (z^{w-1}, \tau^{w-1}, s^w, x^w)$

(iii) Add $\hat{SVI}_w$ to the RMP

(iv) $w \leftarrow w + 1$

Step 2. Solve the RMP, for $(\hat{z}, \hat{\tau})^w$ given $\hat{SVI}_w$. If the RMP is infeasible, then stop, the global minimum has been found and go to Step 3. Else, go to Step 1.

Step 3. Print optimal solution: $(z, \tau, s, x)^* = (z, \tau, s, x)^+$
The correctness of the algorithm is explained here. At iteration \( w \) of the algorithm, the new found protection strategy in Step 2 improves on all those disruption scenarios implied by \( \hat{SVI}_w \). Let \( \hat{SVI}^* \) be the set of SVIs associated with an optimal protection strategy \( (z^*, \tau)^* \). At this point, \( z^* \) must be able to satisfy all of the SVIs in \( \hat{SVI}^* \). However, the RMP will become infeasible at a given iteration \( w_+ \) due to the addition of \( SVI_{w_+} \). As such, no protection strategy exists which is able to meet all the constraints in \( \hat{SVI}_{w_+} \). The optimal protection strategy is then the best out of all the protection strategies found thus far.

4.4 Combining D-Bend and D-SVI: a Hybrid Decomposition (D-H)

The purpose of this section is to devise a hybrid decomposition that captures the best features of D-Bend and D-SVI. Intuitively, D-Bend is likely to require fewer iterations than D-SVI. On the other hand, the RMP in D-Bend is considerably more computationally expensive than the one for D-SVI, as the latter is a feasibility seeking problem. As the results in Section 5 demonstrate, both these conjectures hold true.

We present here a hybrid method D-H, which consists of an easier to solve RMP than the one defined for D-Bend. The aim of D-H is to prove optimality in fewer iterations than D-SVI.

The RMP for D-H has a feasibility region defined by a set of SVIs (as in D-SVI) plus a small number of adaptive Benders optimality cuts (as in D-Bend). The idea is based on the approach of O’Hanley and Church (2011). While the purpose of the SVIs in the RMP is the same as in D-SVI, the inclusion of a few ABCs in the RMP serves to form a lower bound on FRIMT and hence a more robust stopping condition. Benders cuts are introduced into the RMP at only the first few iterations and once added to the RMP, they remain there until the algorithm converges to optimality. Cutting plane SVIs, on the other hand, are introduced into the RMP at every iteration. Finally, D-H converges according to the two stopping criteria established in D-SVI and D-Bend: optimality is proven when the RMP becomes infeasible (according to the stopping rule of D-SVI) or when \( H^{up} - H^{low} < \epsilon \) (according to the stopping criteria of D-Bend).

In our implementation, we include only one adaptive Benders cut (the one from the first iteration). The choice of which ABCs are introduced into the RMP is likely important in reducing the number of iterations. We do not explore here how to further optimize the selection of ABCs in the framework of D-H. A more in-depth investigation on this issue is the subject of future research.
5 Results and analysis

5.1 Computational performance

We tested the computational performance of the proposed approaches (D-Bend, D-SVI and D-H) on an Intel Core 2 Duo T6400 (2 GHz per CPU) with 4 GB of RAM. The algorithms were implemented in C++ using Cplex 12.1 (IBM) callable libraries. The computational tests were conducted on the London Ontario dataset with 150 nodes/demand points (Waters, 1977) and the UKCities dataset with 250 nodes/demand points (Liberatore and Scaparra, 2011).

In our initial testing, we solved each problem for a number of different \( p \) values, (20, 30 and 40) and different values of \( r \) and \( q \) for each value of \( p \). The planning horizon \( T \) was set to 30, the facility recovery time \( G_j \) ranged randomly from 1 to 5 and the reduction in facility recovery time per unit of protection invested on the facility \( m_j \) was set to 1. We set a time limit of 3 hours for each decomposition procedure and an optimality tolerance of 0.1%.

Table 1 displays a comparison among the D-Bend, D-SVI and D-H decomposition algorithms for the London dataset. The objective function value associated with the optimal solution for a particular instance is displayed in column four from the left. The computing times are assorted in three groups of three columns for each algorithm, appearing from left to right as: total computing time, computing time spent on solving RMPs and computing time for solving SPs. The three right most columns indicate the number of iterations elapsed until optimality was reached for D-Bend, D-SVI and D-H, respectively. The performance of D-Bend is clearly inferior to D-SVI and D-H; D-H slightly outperforms on average D-SVI. The average total computing time of D-Bend is 156.60 seconds. D-SVI produced better results with an average time of 76.12 seconds. D-H is the fastest algorithm for these instances with an average total solution time of 25.98 seconds. For this algorithm, the instance with the largest computing time \((p = 40, r = 12, q = 16)\) was solved in less than 200 seconds.

As expected, D-SVI iterates faster than the other decompositions because its RMP does not have an objective function and is terminated at the first feasible node found in a branch and bound (B&B) tree. On the contrary, D-Bend and D-H need to prove optimality for their RMPs, resulting in a considerable increase in overall computing time. As such, D-SVI requires on average 40.78 iterations while D-Bend and D-H only 8.17 and 30.28 iterations, respectively. D-H seems to be an efficient trade-off between the number of elapsed iterations and the complexity of the RMPs.

Given the efficiency of the D-SVI and D-H algorithms, we put them to the test on a larger dataset, the UKCities, and larger values for the \( p, r \) and \( q \) parameters. We report in Table
<table>
<thead>
<tr>
<th>( p )</th>
<th>( r )</th>
<th>( q )</th>
<th>\text{Obj.Func.Val}</th>
<th>\text{Total Time}</th>
<th>\text{RMP Time}</th>
<th>\text{SP Time}</th>
<th>\text{No. Iterations}</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>2</td>
<td>2</td>
<td>3788258.44</td>
<td>0.07</td>
<td>0.11</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>4</td>
<td>3766101.31</td>
<td>0.27</td>
<td>0.21</td>
<td>0.10</td>
<td>0.04</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>4</td>
<td>3869736.46</td>
<td>0.82</td>
<td>0.82</td>
<td>0.50</td>
<td>0.14</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>6</td>
<td>3829484.27</td>
<td>7.09</td>
<td>2.09</td>
<td>0.86</td>
<td>6.12</td>
</tr>
<tr>
<td>20</td>
<td>6</td>
<td>6</td>
<td>3925090.26</td>
<td>2.94</td>
<td>2.08</td>
<td>1.97</td>
<td>2.43</td>
</tr>
<tr>
<td>20</td>
<td>6</td>
<td>8</td>
<td>3884427.83</td>
<td>5.48</td>
<td>5.00</td>
<td>3.31</td>
<td>4.68</td>
</tr>
<tr>
<td>30</td>
<td>3</td>
<td>3</td>
<td>2733274.37</td>
<td>0.23</td>
<td>0.57</td>
<td>0.34</td>
<td>0.10</td>
</tr>
<tr>
<td>30</td>
<td>3</td>
<td>6</td>
<td>2727769.72</td>
<td>2.31</td>
<td>1.50</td>
<td>0.76</td>
<td>1.86</td>
</tr>
<tr>
<td>30</td>
<td>6</td>
<td>6</td>
<td>2804379.84</td>
<td>6.59</td>
<td>4.01</td>
<td>2.71</td>
<td>5.33</td>
</tr>
<tr>
<td>30</td>
<td>6</td>
<td>9</td>
<td>2793365.58</td>
<td>7.53</td>
<td>9.74</td>
<td>4.50</td>
<td>6.89</td>
</tr>
<tr>
<td>30</td>
<td>9</td>
<td>9</td>
<td>2857218.71</td>
<td>60.83</td>
<td>89.33</td>
<td>27.02</td>
<td>54.28</td>
</tr>
<tr>
<td>30</td>
<td>9</td>
<td>12</td>
<td>2842406.13</td>
<td>296.73</td>
<td>143.21</td>
<td>74.32</td>
<td>286.64</td>
</tr>
<tr>
<td>40</td>
<td>4</td>
<td>4</td>
<td>2047099.56</td>
<td>0.22</td>
<td>0.34</td>
<td>0.17</td>
<td>0.09</td>
</tr>
<tr>
<td>40</td>
<td>4</td>
<td>8</td>
<td>2036019.88</td>
<td>1.12</td>
<td>0.90</td>
<td>0.74</td>
<td>0.58</td>
</tr>
<tr>
<td>40</td>
<td>8</td>
<td>8</td>
<td>2108451.59</td>
<td>13.22</td>
<td>16.72</td>
<td>9.01</td>
<td>9.31</td>
</tr>
<tr>
<td>40</td>
<td>8</td>
<td>12</td>
<td>2101418.59</td>
<td>102.96</td>
<td>44.26</td>
<td>35.02</td>
<td>93.02</td>
</tr>
<tr>
<td>40</td>
<td>12</td>
<td>12</td>
<td>2172508.64</td>
<td>478.04</td>
<td>240.08</td>
<td>108.12</td>
<td>431.62</td>
</tr>
<tr>
<td>40</td>
<td>12</td>
<td>16</td>
<td>2149167.83</td>
<td>1832.51</td>
<td>809.23</td>
<td>198.21</td>
<td>1789.17</td>
</tr>
</tbody>
</table>

Average \[\text{Obj.Func.Val} = 156.60, \text{Total Time} = 76.12, \text{RMP Time} = 25.98, \text{SP Time} = 149.59, \text{No. Iterations} = 2.69, 4.81, 7.01, \text{Average} = 73.44, 21.18, 11.11, 40.78, 30.28\]

London dataset, \( T = 30 \), \( G_j = \text{Rand}[1, 5] \) and \( m_j = 1, \forall j \in F \)
the results for D-SVI and D-H with \( p = 20, 40, 60 \) and a range of \( r \) and \( q \) values. From left to right, we display the values of \( p \), \( r \) and \( q \) defining different instances of the problem, the objective function value associated with the optimal solution for a particular instance and, for both D-SVI and D-H, the total computing time, the computing time spent on solving the RMPs and the computing time for solving the SPs. Note that we do not set any time limit here.

<table>
<thead>
<tr>
<th>( p )</th>
<th>( r )</th>
<th>( q )</th>
<th>Obj.Func.Val</th>
<th>Total Time</th>
<th>RMP Time</th>
<th>SP Time</th>
<th>No. Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>2</td>
<td>2</td>
<td>15246068.76</td>
<td>0.17</td>
<td>0.14</td>
<td>0.03</td>
<td>0.1</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>6</td>
<td>15067213.72</td>
<td>0.43</td>
<td>0.26</td>
<td>0.16</td>
<td>0.07</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>12</td>
<td>14898646.16</td>
<td>0.47</td>
<td>0.58</td>
<td>0.16</td>
<td>0.20</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>20</td>
<td>14798687.01</td>
<td>1.06</td>
<td>0.87</td>
<td>0.46</td>
<td>0.33</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>24</td>
<td>7895249.49</td>
<td>0.36</td>
<td>0.47</td>
<td>0.11</td>
<td>0.15</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>28</td>
<td>7500089.39</td>
<td>1.31</td>
<td>1.14</td>
<td>0.56</td>
<td>0.43</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>40</td>
<td>35092169.80</td>
<td>0.12</td>
<td>0.29</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>24</td>
<td>5383097.80</td>
<td>1.05</td>
<td>1.42</td>
<td>0.33</td>
<td>0.19</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>40</td>
<td>1527834.66</td>
<td>3.14</td>
<td>3.36</td>
<td>1.17</td>
<td>1.42</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>2</td>
<td>8119994.40</td>
<td>0.12</td>
<td>0.11</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>12</td>
<td>7958249.49</td>
<td>0.36</td>
<td>0.47</td>
<td>0.11</td>
<td>0.15</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>24</td>
<td>7897596.83</td>
<td>1.31</td>
<td>1.14</td>
<td>0.56</td>
<td>0.43</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>40</td>
<td>7850159.39</td>
<td>2.40</td>
<td>1.89</td>
<td>0.92</td>
<td>0.46</td>
</tr>
<tr>
<td>20</td>
<td>6</td>
<td>2</td>
<td>5383097.80</td>
<td>0.12</td>
<td>0.29</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>20</td>
<td>6</td>
<td>12</td>
<td>5136621.14</td>
<td>1.05</td>
<td>1.42</td>
<td>0.33</td>
<td>0.19</td>
</tr>
<tr>
<td>20</td>
<td>6</td>
<td>24</td>
<td>5090923.73</td>
<td>3.14</td>
<td>3.36</td>
<td>1.17</td>
<td>1.42</td>
</tr>
<tr>
<td>20</td>
<td>6</td>
<td>40</td>
<td>5090923.73</td>
<td>6.00</td>
<td>3.53</td>
<td>1.54</td>
<td>1.17</td>
</tr>
<tr>
<td>20</td>
<td>12</td>
<td>2</td>
<td>9997319.19</td>
<td>4.57</td>
<td>4.18</td>
<td>0.11</td>
<td>0.14</td>
</tr>
<tr>
<td>20</td>
<td>12</td>
<td>6</td>
<td>9008989.14</td>
<td>75.55</td>
<td>67.61</td>
<td>6.01</td>
<td>10.04</td>
</tr>
<tr>
<td>20</td>
<td>12</td>
<td>24</td>
<td>8633928.99</td>
<td>321.43</td>
<td>264.75</td>
<td>137.45</td>
<td>105.28</td>
</tr>
<tr>
<td>20</td>
<td>12</td>
<td>40</td>
<td>836433.48</td>
<td>624.24</td>
<td>488.88</td>
<td>59.88</td>
<td>359.32</td>
</tr>
<tr>
<td>20</td>
<td>12</td>
<td>60</td>
<td>5048052.30</td>
<td>2.05</td>
<td>2.52</td>
<td>0.12</td>
<td>0.20</td>
</tr>
<tr>
<td>20</td>
<td>12</td>
<td>24</td>
<td>5664445.27</td>
<td>19.38</td>
<td>366.15</td>
<td>26.11</td>
<td>168.68</td>
</tr>
<tr>
<td>20</td>
<td>12</td>
<td>40</td>
<td>540945.07</td>
<td>70.84</td>
<td>475.74</td>
<td>163.47</td>
<td>321.04</td>
</tr>
<tr>
<td>20</td>
<td>12</td>
<td>60</td>
<td>5356213.55</td>
<td>12456.18</td>
<td>7680.35</td>
<td>12299.84</td>
<td>7470.46</td>
</tr>
</tbody>
</table>

**Average** | 498.72 | 396.06 | 374.97 | 366.10 | 33.75 | 29.95 | 121.09 | 117.91 |

**UKCities dataset, \( T = 30 \), \( G_j = \text{Rand}[1, 5] \) and \( m_j = 1, \forall j \in F \)**

Looking at the results, D-H still outperforms D-SVI on average but only slightly. We reckon
that this behavior may vary depending upon the dataset under study and that the D-H method may be further improved by studying more efficient ways of selecting the ABCs to be included in the RMP.

5.2 Analyzing the impact of the protection budget on the solution

The strategic investment allocated for protection purposes of a given system may vary according to the perceived vulnerability of the system. The vulnerability of the system, in turn, may depend on several factors: the likelihood that a disruption of a given magnitude occurs, the concentration of critical human and material resources at certain facilities, the level of security already in place at a given facility, etc.

FRIMT serves to identify the optimal protection strategy given a fixed amount of resources. Several key questions that arise are: Is the amount of investment devoted to protection strategies appropriate? Is it cost-efficient? Is the protection budget excessive or does it fall short of efficiently protecting the system? The difficulty associated with answering these questions is due to the fact that there is a huge degree of uncertainty related to potential disruptions. It is, therefore, important to identify values of the protection budget $q$ that are cost-efficient under different circumstances (e.g., different numbers of lost facilities and different recovery times), assuming that one cannot generally anticipate the number of losses in a supply system (Liberatore et al., 2010b), or that the duration of the recovery times may be uncertain.

In this section, we suggest an analysis which may help answering the above questions. The analysis is performed on the UKCities dataset with the following combinations of parameters: $p = 10, r = 2, 4$ and $p = 30, r = 3, 6$, with $T = 30$, $G_j = Unif[1, 5]$ and $m_j = 1$. Fig. 1 shows the percent marginal improvement in reduced overall demand weighted travel distance for an increase of two units in the protection budget.

From this graph we derive some noteworthy insights. First, we observe that for every instance, improvement in system efficiency steadily reduces as $q$ increases. Moreover, it seems that up to a critical value of $q$, call it $q^*$, the reduction in the value of the overall travel distance is much greater than the improvement observed for values of $q$ greater than $q^*$. That is, from $q^*$ onwards, the improvement stabilizes at a more or less constant rate. One may infer from this that investments in protection strategies for values above $q^*$ may not be cost-efficient, or at least not as efficient as investments below $q^*$. Clearly, $q^*$ varies from one system to another. For example, for the UKCities dataset, with $p = 10$ and $r = 4$, the value of $q^*$ is around 10.
Second, the improvement is more pronounced for instances with large values of $r$ and small values of $p$. Intuitively, it is more advantageous to increase the protection efforts when the amount of disruptions increases and in smaller systems as they are usually more vulnerable to disruptions due to fewer built in redundancies. Consequently, the protection of additional components may have a larger impact on these systems. As an example, if we consider a protection budget $q = 6$, it is evident from the Figure 1 that the marginal efficiency improvement is considerably higher in a system with 10 facilities ($p = 10$) rather than in a larger system with 30 facilities ($p = 30$). Third, it is worth noting that from $q^*$ onwards, the marginal improvement is stable and below 1% for all the instances analyzed. These results (not shown) were also observed on other datasets with similar choices of the parameters. This non intuitive finding seems to indicate that the critical protection budget $q^*$ may be cost-efficient for the majority of instances of a given dataset.

In any case, the analysis performed illustrates that by varying the values of the disruption parameters that are likely to be uncertain (e.g., $r$ and $G_j$), one can evaluate the impact of specific disruptions on the system. In turn, this allows a decision planner to choose cost-efficient values of the protection budget $q$ for any particular system.

**Figure 1**: Analysis of the impact of the protection budget on the system’s efficiency

**UKCities dataset, $T = 30$, $G_j = Rand[1, 5]$ and $m_j = 1$, $\forall j \in F$**

6 Conclusions

In this article, the important issue of facility recovery time has been incorporated in a model that identifies the optimal allocation of protection resources in an uncapacitated median
network to hedge against worst-case facility losses. The resulting formulation is a bilevel problem with non-convex regions due to the presence of integer variables in the lower level. It has been shown that classical Benders decomposition can be adapted to solve optimally moderate to large instances of this problem in reasonable time. While Benders decomposition requires a few iterations to converge, the associated restricted master problem (RMP) is hard to solve. A different decomposition algorithm based on a set of problem specific supervariable inequalities (SVI) was also devised. Empirical results indicate that although this decomposition algorithm requires significantly more iterations, it is generally more efficient than Benders decomposition. Finally, a hybrid decomposition approach was presented, which exploits the advantages of the previous two methods. Computational testing on two datasets show that this hybrid decomposition is often even more efficient than a pure SVI based approach. The key advantage of the proposed decomposition approaches is their generality; with suitable modifications, they could be applied to more complex protection problems, including those with capacities on facilities.

Finally, a parametric analysis was conducted to identify cost-efficient values of the protection budget. Empirical results suggest that for a given system, there is a value of the protection budget after which the marginal improvement in efficiency becomes negligible with additional protection resources. This analysis, therefore, highlights the tradeoff between additional protection investments and possible efficiency gains. Ultimately, it helps to identify possible over or under investment of the resources dedicated to protection efforts. This analytical tool is particularly useful when there is a high degree of uncertainty with regard to the type of disruption (e.g., the number of possible losses).

The issue of recovery time is undoubtedly of paramount importance when planning the protection of critical infrastructure systems against catastrophic events. We hope that this model will set the groundwork for the development of other protection models where recovery issues are analyzed in combination with other elements, such as facility capacities, inventory, and multi-tiered/multi-commodity distribution systems.

References


Appendix A. Linearization of DSP(\(\hat{s}\))

Let \(v_{ik} = \mu_{ik}(G_k - \tau_k)s_{ik}\). Also, let \(\Phi_j = \{\min(m_j g_j, G_j) | 0 \leq g_j \leq [G_j/m_j], g_j \in \mathbb{Z}\}, \forall j \in F\), and let \(g_j^l\) be the value of \(g_j\) associated with \(l \in \Phi_j\). The set \(\Phi_j\) represents the possible values of the \(\tau_j\) variable (total reduction in the recovery time of facility \(j\)) while \(g_j\) represents the possible values of the \(z_j\) variables (the amount of protection resources invested in facility \(j\)). An example of how \(\Phi_j\) and \(g_j\) are calculated is as follows: for \(G_j = 10\) and \(m_j = 3\), \([G_j/m_j] = 4\). Then, \(g_j^0 = 0, g_j^3 = 1, g_j^6 = 2, g_j^9 = 3, g_j^{10} = 4\).

Also, let us define the binary variable \(\gamma_j^{g_j^l}\) which takes on one if facility \(j\) is protected with \(g_j^l\) resources, zero otherwise. Then, to enforce the correct behaviour of \(v_{ik}\), the following inequalities are required:

\[
v_{ik} \geq \mu_{ik}(G_k - \tau_k)s_{ik} - M_{ik}(G_k - \tau_k)(1 - \gamma_j^{g_j^l}), \forall i \in N, k \leq |F| - 1, l \in \Phi_{ik}\]  
(37)

\[
\tau_j = \sum_{l \in \Phi_j} l\gamma_j^{g_j^l}, \forall j \in F\]  
(38)

\[
\sum_{l \in \Phi_j} \gamma_j^{g_j^l} = 1, \forall j \in F\]  
(39)

The \(M_{ik}\) in the linearization constraints (37) are any sufficiently large scalars. An efficient choice for their values, which tightens the formulation, is discussed in Appendix C. Constraints (38) and (39) ensure that only one “protection level” is given to facility \(j\) (including the level of no protection, \(g_j = 0\)). Constraints (37) require that if \(\gamma_j^{g_j^l} = 0\), then \(v_{ik}\) is not bounded from below. Otherwise if \(\gamma_j^{g_j^l} = 1\), then due to (38), which fixes \(l\) for a fixed value of \(\tau_{ik}\), and due to the minimization objective in (30), it holds that \(v_{ik} = \mu_{ik}(G_k - \tau_k)s_{ik}\).

Appendix B. The RMP Formulation

We generate an ABC for each \(s^b \in \hat{W}_w\) and append it to the RMP. This implies that the RMP has \(|\hat{W}_w|\) vector variables \(\mu\), \(\beta\) and \(v\). Let the decision variable \(H^{low} \in \mathbb{R}_+\) be the objective function value of the RMP. Then, at iteration \(w\) of D-Bend, the RMP is:

[RMP-D-Bend]
\[
\min_{z,\mu,\nu,\beta,\gamma, H^{low}} \quad H^{low} \\
\text{s.t.} \quad \sum_{j} z_j \leq q \quad (40) \\
z_j \leq \left[ G_j/m_j \right], \forall j \in F \quad (41) \\
\tau_j \leq G_j, \forall j \in F \quad (42) \\
\tau_j \leq m_j z_j, \forall j \in F \quad (43) \\
z_j \in \mathbb{Z}_+, \forall j \in F \quad (44) \\
\tau_j \in \mathbb{R}_+, \forall j \in F \quad (45) \\
H^{low} \geq \sum_{i} \left( \sum_{k} v_{iik}^b + T \beta_k^b \right), \forall b = 1, ..., w \quad (46) \\
\sum_{l} \mu_{ii}^b + \beta_k^b \geq h_i d_{ii}, \forall i \in N, 1 < k \leq |F|, b = 1, ..., w \quad (47) \\
\beta_k^b \geq h_i d_{ii}, \forall i \in N, b = 1, ..., w \quad (48) \\
v_{iik}^b \geq \mu_{ii}^b (G_i - l)s_{ik}^b - M(G_i - l)s_{ik}^b (1 - \gamma_{iik}^b), \forall i \in N, 1 < k \leq |F| - 1, \quad \forall i \in N, 1 < k \leq |F| - 1, l \in \Phi_{ik}, b = 1, ..., w \quad (49) \\
\tau_j = \sum_{l} \gamma_{jkl}, \forall j \in F \quad (50) \\
\sum_{l} \gamma_{jkl} = 1, \forall j \in F \quad (51) \\
\mu_{ii}^b \in \mathbb{R}_+, \forall i \in N, 1 \leq k \leq |F| - 1, b = 1, ..., w \quad (52) \\
\beta_k^b \in \mathbb{R}_+, \forall i \in N, b = 1, ..., w \quad (53) \\
v_{iik}^b \in \mathbb{R}_+, \forall i \in N, 1 \leq k \leq |F| - 1, b = 1, ..., w \quad (54) \\
\gamma_{jkl} \in \{0,1\}, \forall j \in F, l \in \Phi_j \quad (55) \\
\gamma_{jkl} \in \{0,1\}, \forall j \in F, l \in \Phi_j \quad (56) 
\]

Note that RMP-D-Bend is obtained by transforming a bilevel program with multiple lower level programs into a single-level program by dualizing each lower level subproblem. That is, the RMP is initially formulated as a bilevel problem, where the upper level program, \(40)-(46)\), includes integer variables \(\text{the protection decision variables}\) and a set of \(|\hat{W}_w|\) lower level programs. Each lower level program corresponds to \(47)-(56)\) for \(s^b \in \hat{W}_w\), which is an optimization problem over continuous variables that correctly enforces the assignment of customers to facilities given \(s^b\) and protection strategy \(\tau\) established by the upper level program. By taking the dual of the lower level problems using \(37)-(39)\) and moving them into the upper level, the resulting RMP-D-Bend formulation is obtained. In summary, \(47)-(56)\) represent the set of adaptive Benders cuts obtained by dualization of each SP.
Appendix C. Variable bounding and reduction for D-Bend

The dual variables $\mu_{iik}^b$ and $v_{iik}^b$ are defined only for $1 \leq k < r + 1$, according to constraint (28) and the reduction rule for the assignment variables $x$. The following lemmas show how the number of dual variables can be further reduced and how valid upper bounds can be identified for the remaining variables.

Let $i_{k0}$ be the closest facility to customer $i$ that has not been interdicted. Then for each interdiction pattern $s^b \in \hat{W}$, we can reduce the number of $\mu^b$ and $v^b$ variables in the following way.

**Lemma 1:** For each $k$ such that $k^0 < k \leq r$, the dual variables $\mu_{iik}^b$ can be set equal to zero.

**Proof:** By definition, facility $i_{k0}$ is not interdicted (i.e., $s^b_{i_{k0}} = 0$). Since in the objective function (30) of each adaptive bender cut (the DSP$(\hat{s})$ programs before linearization), the dual variables $\mu$ are multiplied by the variables $s$, the variables $\mu_{iik}^b$, can be set as high as desired without affecting the objective function value of the program. Consequently, the value of $\sum_{k=1}^{k^0} \mu_{iik}^b$ always meets constraints (31) (or constraints (48) in the linearized formulation) and the dual variables corresponding to facilities further away than $i_{k0}$ can be eliminated. $\square$

**Lemma 2:** For each $k$ such that $k^0 < k \leq r$, the dual variables $v_{iik}^b$ can be set equal to zero.

**Proof:** From Lemma 1, $\mu_{iik}^b = 0$, $\forall k : k^0 < k \leq r$. Therefore, constraints (50) reduce to $v_{iik}^b \geq \mu_{iik}^b - M(G_{ik} - l)s^b_{i_{k0}}(1 - \gamma_{i_{k0}}^b s^b_{i_{k0}})$, $\forall s^b \in \hat{W}, i \in N, k^0 < k \leq r, l \in \Phi_{ik}$. Since the objective function of RMP-D-Bend minimizes the value of the $v$ variables, these must be zero and can be eliminated. $\square$

**Lemma 3:** The dual variables $v_{iik0}^b$ can be set equal to zero.

**Proof:** This follows directly from constraints (50) when $s^b_{i_{k0}} = 0$ and the fact that the objective function of RMP-D-Bend seeks to minimize the value of the variables $v$.

For the remaining $\mu^b$ and $v^b$ variables, as well as the $\beta^b$ variables, valid upper bounds can be established in the following way.

**Lemma 4:** A valid upper bound for the variables $\mu_{iik0}^b$ is $h_i \left( d_{ii[r+1]} - d_{ii1} \right)$.

**Proof:** From constraints (48), we have $\mu_{iik0}^b \geq h_i d_{ii0[r+1]} - (\beta_i^b + \sum_{l<k^0} \mu_{iil}^b)$. Further, since the objective function of MP-D-Bend seeks to minimize the value of the variables $\mu$ and $\beta$, we have that $\mu_{iik0}^b \leq h_i d_{ii0[r+1]} - (\beta_i^b + \sum_{l<k^0} \mu_{iil}^b) \leq h_i d_{ii[r+1]} - d_{ii1} \leq h_i (d_{ii[r+1]} - d_{ii1})$. $\square$

**Lemma 5:** For each $k$ such that $1 \leq k < k^0$, a valid upper bound for $\mu_{iik}$ is $h_i \left( d_{iik+1} - d_{ii1} \right)$.

**Proof:** The proof is similar to the one for Lemma 4. That is, from constraints (48) and the
fact that the objective function of RMP-D-Bend must be minimized, an upper bound of $\mu_{iik}^b$ is

$$\mu_{iik}^b \leq h_i d_{iik+1} - (\beta_i + \sum_{l<k} \mu_{iik}^b) \leq h_i (d_{iik+1} - d_{i1}) \quad \Box$$

**Lemma 6**: For each $k$ such that $1 \leq k < k^0$, a valid upper bound for $v_{iik}^b$ is $h_i (d_{iik+1} - d_{i1}) G_{iik} s_{iik}^b$.

**Proof**: From constraints (50), we have that $v_{iik}^b \geq \mu_{iik}^b G_{iik} s_{iik}^b$ when $\gamma_{iik} = 1$. Further, since the objective function of RMP-D-Bend minimizes the value of the $v$ variables and the maximum value of $\mu_{iik}^b$ is $h_i (d_{iik+1} - d_{i1})$, as stated in Lemma 5, then $v_{iik}^b \leq h_i (d_{iik+1} - d_{i1}) G_{iik} s_{iik}^b$. \quad \Box

**Lemma 7**: A valid upper bound for $\beta_i$ is $h_id_{i(i+1)}$.

**Proof**: This follows from constraints (48) when the $\mu$ variables take on zero values. \quad \Box

The following lemma sets efficient values for the scalars $M_{iik}^b$ so as to tighten the linearized formulation RMP-D-Bend.

**Lemma 8**: The constants $M_{iik}^b$ in RMP-D-Bend can be set equal to the upper bounds of the variables $\mu_{iik}^b$ established in Lemmas 4 and 5.

**Proof**: From (50), the maximum value of $M_{iik}^b$ occurs when $\gamma_{iik} = 0$. In this case, the variables $v_{iik}^b$ are pushed to zero given the minimization of the objective function for RMP-D-Bend. Therefore, we have that $M_{iik}^b \geq \mu_{iik}^b$ and the bounds previously established for the variables $\mu_{iik}^b$ are valid upper bounds for the constants $M_{iik}^b$. \quad \Box

Finally, we propose a reduction rule for the protection variables $z_j$.

**Lemma 9**: At iteration $w$, a variable $z_j$ can be set to zero if facility $j$ does not appear in any of the $w$ previous interdiction strategies. (Note that this rule is only valid at iteration $w$, i.e. in subsequent iterations this setting is removed unless the facility meets the requirement again).

**Proof**: If facility $j$ does not appear in any interdiction pattern seen so far, there is no gain for RMP-D-Bend in protecting this facility. \quad \Box