Model and a solution algorithm for the dynamic resource allocation problem for large-scale transportation network evacuation

Xiaozheng He, Hong Zheng, Srinivas Peeta

Abstract

Allocating movable resources dynamically enables evacuation management agencies to improve evacuation system performance in both the spatial and temporal dimensions. This study proposes a mixed integer linear program (MILP) model to address the dynamic resource allocation problem for transportation evacuation planning on large-scale networks. The proposed model is built on the earliest arrival flow formulation that significantly reduces problem size. A set of binary variables, specifically, the beginning and the ending time of resource allocation at a location, enable a strong formulation with tight constraints. A solution algorithm is developed to solve for an optimal solution on large-scale network applications by adopting Benders decomposition. In this algorithm, the MILP model is decomposed into two sub-problems. The first sub-problem, called the restricted master problem, identifies a feasible dynamic resource allocation plan. The second sub-problem, called the auxiliary problem, models dynamic traffic assignment in the evacuation network given a resource allocation plan. A numerical study is performed on the Dallas-Fort Worth network. The results show that the Benders decomposition algorithm can solve an optimal solution efficiently on a large-scale network.

Keywords: Evacuation planning; resource allocation; mixed integer linear program; Benders decomposition

1. Introduction

Efficient transportation of evacuees during emergencies has long been recognized as a challenging issue. Effective evacuation plans play an important role in improving the evacuation throughput, and reduce clearance time and property losses. In the evacuation literature, some studies focus on modeling traffic flow characteristics and dynamics during evacuation (see reviews in Murray-Tuite and Wolshon, 2013 and Dixit and Wolshon, 2014). Other studies focus on improving the ground transportation performance during an evacuation through network design, mainly including: (i) network topological design, such as expanding link capacity (Patil and Ukkusuri 2007; Ng and Waller 2009), planning lane reversal strategies (so-called contraflow or counterflow) (Tuydes and Ziliaskopoulos 2006; Kalafatas and Peeta 2009; Xie et al. 2010), or enhancing link functionality reliability (Peeta et al. 2010); (ii) intersection design, such as reducing crossing and merging conflicts by restricting the right-of-way at intersections (Cova and Johnson 2003; Xie and Turnquist 2011); and (iii) traffic management strategies, such as investigating shelter location and the associated routing guidance to the shelter (Sherali et al. 1991; Kongsomsaksakul et al. 2005;
Ng et al. 2010), signal timing to favor evacuation traffic (Lo 2001; Chen et al. 2007), stage-based traffic flow control using behavior-consistent information (Hsu and Peeta, 2014a, 2014b), and law enforcement personnel deployment strategy (Jabari et al. 2012). These strategies relevant to evacuation network design and management primarily deal with the optimal distribution of limited resources, such as budget, personnel, and traffic control devices, in the evacuation process. Among most existing evacuation network design studies, resource locations (or design patterns) are fixed, namely, they do not change over time during the evacuation process. Resource allocation in a dynamic model has been addressed in large-scale disasters, for instance, earthquake (Bakuli and Smith 1996; Fiedrich et al. 2000). Because time and the quantity of the resources are limited, emergency management agencies need to find an optimal schedule for assigning resources in space and time to the threatened areas to enable a timely evacuation.

This paper investigates the dynamic resource allocation of well-trained personnel and movable devices to different intersections to route the evacuation traffic as efficiently as possible through these intersections. In this context, the personnel could be used to control traffic through an intersection to preclude, for example, turning movements that can reduce capacity under saturated conditions. Similarly, movable message signs could be used to disseminate real-time route guidance information towards safe zones. Allocating these movable resources dynamically can enable management agencies to further improve the system performance in the evacuation process. The study seeks to generate movable resources allocation plans to support evacuation planning, especially on large-scale networks. When a disaster occurs, the management agencies can deploy the plans generated off-line based on the unfolding traffic conditions.

From a modeling standpoint, the dynamic resource allocation (DRA) problem has some similarity to the machine scheduling problem in the operations research domain, in which a number of known jobs (i.e., resources in our context) need to be assigned to variant machines (i.e., locations in our context), such that the obtained schedule leads to minimum costs (Thomas and Salhi 1998). As the same resource can be dynamically allocated to different locations, the DRA model is characterized by the job-shop problem where a job (one resource/personnel) can be assigned to different machines (locations).

To establish a foundation for the DRA implementation, He and Peeta (2014) recently proposed a mixed integer linear programming (MILP) formulation for the DRA problem to support evacuation. It determines an optimal plan that dynamically assigns a limited number of moveable resources (e.g., trained personnel) to a set of intersections such that the total system travel time is minimized. The total system travel time includes the waiting time at origins and the travel times through the network to the safe zones. Incorporating congestion, or traffic flow dynamics using concepts of traffic flow theory, is a central component in the constraint set of the problem formulation. In this context, He and Peeta’s DRA model is built upon the Cell Transmission Model (CTM) (Daganzo 1994, 1995) to describe the traffic flow dynamics. In addition, the model includes a set of spatiotemporal constraints to enable realism in practice; for example, the time required to reallocate the resource from one location to another is subject to the feasible traversal time, the allocated resource at one location is subject to the minimum service time, and the gap between successive allocations of resources to the same location is subject to a minimum time span.

The major barrier to applying He and Peeta’s MILP to large-scale evacuation management lies in the high computational cost due to the following analytical features. First, the embedded CTM discretizes the network into small pieces of cells and the time span of interest into small time steps. The model contains a large number of variables because each link is divided into cells at each time step. Hence, a network of even a modest size will create millions of variables. This makes the MILP difficult to solve especially when the network size and evacuation planning time period are large. Second, the existence of traffic backward propagation constraint in the CTM involves a parameter that is not 1, 0, or -1, and thus precludes the time-expanded network structure which is preferable for developing efficient network algorithms. As illustrated by Kalafatas and Peeta (2007), a model that maintains a time-expanded network structure can be solved more efficiently due to the inherent total unimodularity and acyclic properties. Third, the embedded CTM permits traffic holding at intermediate nodes that is unrealistic during evacuation. Fourth, the binary decision variables which determine the allocation of the moveable resources make the problem NP-hard. The number of binary variables would increase significantly if one attempts to increase fidelity of traffic dynamics by dividing cells at a smaller dimension. Fifth, He and Peeta (2014) use a heuristic algorithm to identify an approximate solution that does not guarantee optimality unlike the proposed study. Hence, there is a key need for a DRA model and an efficient solution algorithm to facilitate practical applications, especially for evacuation planning for large-scale networks.
This study addresses the DRA problem especially for large-scale transportation network evacuation in two aspects. First, from a modeling standpoint, we adapt the earliest arrival flow (EAF) model instead of the CTM to describe the system-optimun evacuation traffic flow pattern. The main difference between the EAF model and the CTM is the choice of how to model dynamic traffic flows in the constraint sets of the DRA problem. The CTM describes realistic traffic flow dynamics but destroys the graph-theoretic properties and makes the problem much harder to solve. Recently, Zheng et al. (2015) illustrate that the traffic flow component does not matter much for certain evacuation objectives, especially if route advisories and guidance can manage flows. They show that the EAF and the system optimum dynamic traffic assignment (SO-DTA) encapsulating CTM have the same total system travel time. Another advantage of using the EAF model is that the evacuation flow can be modeled on a much simpler link-node time-expanded network (without division of cells) without compromising the optimal solution. In addition, traffic holding can be precluded in the EAF model.

Second, to improve the computational efficiency, a solution algorithm based on the Benders decomposition scheme (Benders, 1962) is developed for the DRA problem. Decomposition is a common technique to solve large-scale optimization problems. In this paper, the DRA model has a separable structure, which leverages the benefits of using Benders decomposition to solve the problem efficiently. The idea is to decompose the DRA problem into two sub-problems that are easier to solve. The restricted master problem assigns the moveable resources dynamically to the network and ensures assignment feasibility. The auxiliary problem solves for the evacuation flows in the network with the fixed resource allocation plan. In addition, the auxiliary problem can be solved by one call of the EAF problem efficiently. The dual variables of the auxiliary problem generate a Benders cut feedback to the restricted master problem. The master and auxiliary problems are solved iteratively until the convergence criterion is satisfied and the optimal solution is identified for the DRA problem. Therefore, the efficiency of solving the master and auxiliary problems is the key to achieving high computational efficiency.

The remainder of the paper is organized as follows. The next section presents preliminaries including notation and assumptions, a brief overview of the EAF problem, and the Benders decomposition scheme. Section 3 describes the details of the proposed MILP for the DRA problem. Section 4 develops a solution algorithm for the DRA problem based on the Benders decomposition scheme. A numerical experiment is shown in Section 5 to illustrate the effectiveness of the proposed model and its solution algorithm on a large-scale network. The final section summarizes this study and identifies potential venues for future research.

2. Preliminaries

This section introduces the notation and assumptions of the formulated DRA model, a brief review of the earliest arrival flow problem, as well as the Benders decomposition scheme that will be used to develop the solution algorithm for the DRA for large-scale network applications.

2.1. Notation and assumptions

Consider the DRA problem on a transportation network denoted by $\mathcal{G}(\mathcal{N}, \mathcal{L})$. $\mathcal{N}$ represents a set of nodes and $\mathcal{L}$ denotes a set of arcs, where the nodes correspond to intersections, or merging and diverging points; and the arcs correspond to road links. A subset of nodes $\hat{\mathcal{N}} \subseteq \mathcal{N}$ with cardinality $|\hat{\mathcal{N}}| = M$ represents $M$ candidate locations for deploying moveable resources. Denote by $\mathcal{P}$ a set of resources to be assigned dynamically in the network. The total number of moveable resources is $|\mathcal{P}| = P < M$. The evacuation time horizon is divided into a set of discrete time steps $T = \{1, ..., T\}$. Let $\mathcal{S}^s \subset \mathcal{N}$ represents the set of origins and $s \in \mathcal{N}$ represents a virtual single destination, also known as the super sink in evacuation studies (Chiu et al., 2007), which connects all destinations (safe zones). Note that this paper focuses on the single destination network structure. This, however, does not restrict the evacuation problem since multiple destinations could, without loss of generality, be connected to the virtual super sink so that the single destination network structure applies.

The notation used includes: (i) $i, j, k \in \mathcal{N}$ as indices for nodes, (ii) $(i, j) \in \mathcal{L}$ or $ij \in \mathcal{L}$ to represent a specific link with tail node $i$ and head node $j$, (iii) $p \in \mathcal{P}$ as an index for moveable resources to be assigned to the evacuation network, and (iv) $t, T, \theta \in T$ as indices for time steps. Superscript $'$ is used to represent transpose of a vector or matrix. Additional notation will be introduced when necessary.

The DRA problem in this study has the following assumptions:
Assumption 1. At most one moveable resource can be assigned to an intersection once throughout the evacuation process.
Assumption 2. The minimum time a moveable resource is assigned to a location is at least $H > 0$ consecutive time steps.
Assumption 3. The resource transition time from location $i$ to $j$ is subject to a constant value $c_{ij}$ throughout the evacuation process.

From a modeling standpoint, Assumption 1 assists to formulate the DRA problem with a linear constraint set, which makes the model tractable. For no-notice evacuation, as all evacuation demand is loaded into the network at the beginning time, the assigned moveable resource should stay at the same bottleneck before evacuees get cleared. If a moveable resource is assigned to another location after serving one location without demand being cleared, it would result in throughput loss during the transition time compared to staying at the same location. Thus, Assumption 1 is reasonable in this context.

Assumption 2 aims to reduce the frequency of resource reallocation during evacuation. As discussed in He and Peeta (2014), frequent changes in the location of a resource can reduce evacuation efficiency, because moving a resource from one location to another impacts the surrounding evacuation traffic. However, it is theoretically possible that achieving a global system optimum may require adjusting resource locations frequently. In this context, Assumption 2 can help emergency management agencies preclude DRA solutions with impractical deployments, and focus on those assigning limited resources to the most critical intersections without additional operational complexity.

Assumption 3 indicates that a transition time is necessary to move a resource from one location to another. The transition time includes the resource travel time in the network as well as the redeployment time at the new location. In practice, the transition time between two locations can be affected by the traffic conditions. However, incorporating time-dependent travel time increases model complexity and computational burden. Further, as moveable resources are usually transferred using emergency vehicles that have the right-of-way throughout the evacuation process, the transition time between two specific locations can be approximated as having a fixed value during evacuation.

2.2. Earliest arrival flow formulation

In He and Peeta (2014), the CTM is deployed to model traffic flow dynamics during evacuation. It is usually believed that the CTM can more realistically model flow dynamics and queue spillback. However, the CTM requires the division of a link into small cells and the time horizon into small time steps, resulting in a prohibitively large number of variables.

Zheng et al. (2015) demonstrate that the CTM constraints could be relaxed in the SO-DTA on a single destination network. They show that SO-DTA without CTM constraints is theoretically one optimal solution of the SO-DTA problem embedded with the CTM. Their findings also apply for the evacuation problem, which is characterized by the objective of minimizing the total system travel time on a single destination network. The simplified problem is called the EAF problem, which finds a dynamic flow with the earliest arrival pattern on the original node-link network without the division of cells.

Given a time horizon $T$ and a dynamic network $G = \left( N, L, q, \eta, T \right)$, where each arc $(i, j) \in L$ is associated with a capacity $q_{ij}$ and a traversal time $\eta_{ij}$, a flow over time is called an EAF if the amount of flow that has already reached the sink until time $\tau$ is maximal simultaneously for all possible time points $0 \leq \tau \leq T$. Let $v_{\text{max}}(\tau)$ be the value of the maximum dynamic flow that can be sent from all sources to the sink until time $\tau$; an EAF achieves the value of $v_{\text{max}}(\tau)$ at any time point $\tau$ such that $0 \leq \tau \leq T$.

The amount of exiting flow is maximal at every time point in an EAF. Fig. 1 illustrates the arrival pattern of an EAF. If there is a single origin, the exiting flow at the destination node is convex; if there are multiple origins, the exiting flow is not convex, i.e., the flow rate could drop when an origin depletes its demand. For example, Fig. 1 demonstrates that sources $S_1$, $S_2$, and $S_3$ become empty at times $\tau_1$, $\tau_2$, and $\tau_3$, respectively, characterized by a declining arrival rate.
Fig. 1. Arrival pattern of an earliest arrival flow.

Gale (1959) and Philpott (1990) showed, in discrete and continuous time respectively, the existence of an EAF on a single-source single-sink network. It should be noted that an EAF may not exist on a network with multiple sinks. Fleischer (2001) presented an instance on a network with two sinks where an EAF does not exist. On a single-sink network with multiple sources, an EAF does exist if the network parameters are pre-defined (Fleischer and Skutella 2007).

In this study, the EAF of Zheng et al. (2015) is employed. The formulation of the EAF problem is represented by Eqs. (1) – (6).

EAF: \[ \min \ z = \sum_{(j,s) \in L} \int_0^T \tau \cdot x_{js} (\tau) \, d\tau \]  
subject to:

\[ \int_0^T x_{ji} (\tau) \, d\tau = b_i \quad \forall i \in S^+ \]  
\[ \sum_{(j,s) \in L} \int_0^T x_{js} (\tau - \nu_{js}) \, d\tau = \sum_{i \in S^+} b_i \]  
\[ \sum_{(j,s) \in L} \int_0^T x_{js} (\tau - \nu_{js}) \, d\tau - \sum_{(i,) \in \mathcal{L}} \int_0^T x_{ij} (\tau) \, d\tau \geq 0 \quad \forall i \in N \setminus \{S^+ \cup S\}, \forall \theta \in \mathcal{T} \]  
\[ \sum_{(i,) \in \mathcal{L}} \int_0^T x_{ij} (\tau - \nu_{ji}) \, d\tau - \sum_{(i,) \in \mathcal{L}} \int_0^T x_{ij} (\tau) \, d\tau = 0 \quad \forall i \in N \setminus \{S^+ \cup S\} \]  
\[ 0 \leq x_{ij} (\tau) \leq q_{ij}, \forall (i,j) \in \mathcal{L}, \tau \in \mathcal{T}. \]
2.3. Benders decomposition

The Benders decomposition scheme is well known for solving stochastic programming and mixed integer programming problems, which typically have a separable structure. Benders decomposition has been widely applied to problems in many domains such as electricity supply investment (e.g., Anderson 1972), industrial allocation system design (e.g., Geoffrion and Graves 1974), water resources management (e.g., Cai et al. 2001), and hydrothermal coordination problems (e.g., Cerisola and Ramos 2002).

The Benders decomposition scheme has long been applied to problems in transportation domain. Florian et al. (1976) applied this scheme to locomotive scheduling problem. Richardson (1976) developed a solution algorithm based on Benders decomposition to solve aircraft routing problem. Fisher and Jaikumar (1981) applied this scheme to design a heuristic method for vehicle routing problem. Marufuzzaman et al. (2014) applied the Benders decomposition method to design a hub-and-spoke network for a bio fuel supply chain system.

Consider an optimization problem with the following structure:

$$\min \left\{ |z + f(y)| : \begin{array}{l} Ax + F(y) \leq g, y \in Y, x \geq 0 \end{array} \right\}$$

where decision variables $y$ are binary and thus are difficult to solve. The main idea behind Benders decomposition is to solve an easier problem over $x$ by temporarily fixing $y$, then use the parameterized solution $x[y]$ and the duality property to improve the estimate of the optimal value of $y$. The original Benders decomposition scheme requires that the remaining problem is a linear program after $y$ are fixed, such that duality theorem in linear programming can be used to generate cutting planes to solve for the difficult problem over variables $y$. Later, Geoffrion (1972) developed a generalized Benders decomposition scheme, in which the sub-problem under the fixed $y$ can be nonlinear.

For an optimization problem with the structure as (7), Benders decomposition contains the following four steps (Bertsimas and Tsitsiklis, 1997):

**Step 0:** Identify a feasible dynamic resource allocation solution $y$ by solving the restricted master problem:

$$\min \{ z + f(y) | y \in Y \}.$$  

Denote $z^*$ the optimal value of $z$.

**Step 1:** For the currently fixed $y$, solve the dual problem of $\min \{ w'x + f(y) | Ax \leq g - F(y), x \geq 0 \}$, i.e.:

$$\max \left\{ \left( g - F(y) \right)' u : A'u \leq w, u \geq 0 \right\}.$$  

**Step 2:** If the dual problem is unbounded, then generate a Benders feasibility cut:

$$\left( g - F(y) \right)' u \leq 0.$$  

Add the cut (9) into the restricted master problem. Go to Step 3.

If the dual problem is bounded and its objective function value is greater than $z^*$, then generate a Benders optimality cut:

$$\left( g - F(y^*) \right)' u^* \leq z.$$  

Add the cut (10) into the restricted master problem. Go to Step 3.

If the dual problem is bounded and its objective function value is equal to the current value of $z^*$, the problem is solved to optimality, and the process is stopped.

If the dual problem is infeasible, it indicates that the primal problem is either infeasible or unbounded, and the process is stopped.

**Step 3:** Solve the restricted master problem:

$$\min \left\{ z + f(y) | \begin{array}{l} z \geq \left( g - F(y) \right)' u, 0 \geq \left[ g - F(y) \right]' u, y \in Y \end{array} \right\}.$$  

Update variables $y$ and $z^*$. Go to Step 1.
As Benders decomposition requires to solve the auxiliary problem (8) and the restricted master problem (11) iteratively, its computational efficiency depends heavily on the effectiveness of the Benders cuts produced by the auxiliary problem (8) and the efficiency of solving the restricted master problem. To enhance the computational efficiency, additional cuts may be introduced into the restricted master problem (11) to identify an optimal solution more quickly.

3. Formulating the dynamic resource allocation problem

In this section, a mixed integer linear programming (MILP) model is developed for the DRA problem under no-notice evacuation. The proposed model determines the schedule of assigning limited number of moveable resources to appropriate intersections during the evacuation time horizon. It employs the EAF (Zheng et al. 2015) to determine the evacuation routing. Binary decision variables are introduced to determine the locations where each moveable resource is assigned to and the duration that an assigned resource should stay at a specific location. All the moveable resources are assumed to route the vehicles through the intersections without crossing-conflicts, as discussed in Nassir et al. (2014). The rest of this section will first present the decision variables used in the model, and then construct the objective function and constraints in the MILP model, followed by a summary of the analytical properties of the proposed model.

3.1. Decision variables

Let $y_{i_p}$ be binary decision variables in the proposed model. If resource $p$ is assigned to location $i$ at time step $\tau$, $y_{i_p}^\tau = 1$; otherwise, $y_{i_p}^\tau = 0$. In addition, let $S_{i_p}^\tau$ and $E_{i_p}^\tau$ be binary auxiliary variables. If $\tau$ is the first time step in which resource $p$ is assigned to location $i$, $S_{i_p}^\tau = 1$; otherwise, $S_{i_p}^\tau = 0$. And if $\tau$ is the last time step in which resource $p$ is assigned to location $i$, $E_{i_p}^\tau = 1$; otherwise, $E_{i_p}^\tau = 0$.

In the MILP model developed in He and Peeta (2014), only binary variables $y_{i_p}^\tau$ are used as decision variables associated with the resource assignment. As binary variables $y_{i_p}^\tau$ indicate the assignment status at individual time steps and do not account for the continuity of the assignment, a large number of 1s need to be determined by the program. As a result, modeling the DRA problem purely based on binary variables $y_{i_p}^\tau$ induces a high computational complexity due to the inherent combinatorial feature.

By contrast, the proposed model introduces the start and end time steps of the resource assignment schedule. As the model focuses on determining the start and end time steps of an assignment schedule for each resource at each location, additional constraints are added into the model formulation, which can provide tighter cuts in developing solution algorithms. This is based on the observation that, for given resource $p$ and location $i$, once the start time step $\bar{\tau}$ and end time step $\tau$ are determined, i.e., $S_{i_p}^{\bar{\tau}} = 1$ and $E_{i_p}^\tau = 1$, then for any time step $\bar{\tau} \leq \tau \leq \tau$, $y_{i_p}^\tau = 1$.

Another set of decision variables, $x_{i_p}^{ij}$, represent the evacuation flows on link $(i, j)$ at time point $\tau$. When the dynamic resource allocation plan is given, the dynamic evacuation flows $x_{ij}^\tau$ can be determined by solving the EAF (1)-(6). Compared with the model in He and Peeta (2014) built upon the CTM, the problem size is much smaller as the EAF does not need to divide the link into cells. More importantly, the EAF maintains a time-expanded network structure, which assists in developing efficient algorithms for large-scale networks.

3.2. Objective function

The objective of the proposed model is to minimize the total system travel time of all evacuees in the evacuation process, which is a commonly used objective of the SO-DTA. The total system travel time can be computed by the summation of vehicles assigned on the link multiplied by the link traversal time. This SO-DTA objective can be transformed into minimizing the turnstile cost. The turnstile cost among links in this setting is defined as non-zero if and only if the link is connected to the single destination node, or the super sink node, and is zero for all other links. The turnstile cost measures the moment of time when evacuees reach their destinations. Hence, minimizing total travel time where the time function is represented by the turnstile cost – also known as the min-turnstile cost problem – leads to addressing the SO-DTA where the cost function is uniform in the network. The discrete-time formulation of the min-turnstile cost can be represented as:
This case is impossible indicates that the resource allocation is completed before time allocated.

Equality resource has been allocated during the evacuation planning time.

For either inequality and resource set and the evacuation process, due to Assumption 1. Therefore, the summation of the binary decision variables and which facilitate the development of an efficient solution algorithm.

3.3.1. Resource assignment constraints

The first set of constraints states that at most one resource can be assigned to an intersection throughout the evacuation process. Each set of constraints corresponds to an assumption presented in Section 2.1. These constraints are mainly built upon the binary decision variables and which ensure that the clearance time is minimized.

Equality ensures that must have the same value as if both are equal to 0, it represents that no resource has been allocated; otherwise, when both are equal to 1, it implies that resource has been allocated at location .

If the start and end time step variables, i.e. and , respectively, are determined, then binary variables can be specified by:

There are four possibilities: (1) ; it indicates that resource has not been allocated at location by time step ; and thus ; (2) ; it indicates that resource has been allocated to location by time step and is still located there, and thus ; (3) ; it indicates that the resource allocation is completed before time step . Thus ; (4) . Equation (17) maintains the linearity of the model, but cannot ensure that only one unit of resource can be assigned to one location at any time. Therefore, the following inequalities are needed:
\[ \sum_{p=1}^{P} y_{ip}^\tau \leq 1 \quad \forall i \in \hat{N}, \tau \in T, \] (18)

and

\[ \sum_{i=1}^{M} y_{ip}^\tau \leq 1 \quad \forall p \in P, \tau \in T. \] (19)

Inequality (18) indicates that at most one resource can be assigned to an intersection at any time. Inequality (19) ensures that a resource would not be assigned to more than one location at any time.

### 3.3.2. Minimum time that resources are assigned to a location

The second set of constraints seeks to reduce the resource reallocation frequency, as discussed in Section 2.1. Based on Assumption 2, the minimum time for a moveable resource to be assigned to a location is expected to be at least \( H > 0 \) consecutive time steps. Using the start time and end time step variables, this constraint can be formulated as follows:

\[ \sum_{t=1}^{T} E_{ip}^\tau \cdot \tau - \sum_{t=1}^{T} S_{ip}^\tau \cdot \tau \geq H \cdot \sum_{t=1}^{T} S_{ip}^\tau \quad \forall i \in \hat{N}, p \in P. \] (20)

If resource \( p \) is assigned to location \( i \), only one variable \( S_{ip}^\tau \) is equal to one among all time steps \( \tau \in T \) due to inequality (14). Assume that \( \bar{\tau} \) is the first time step that resource \( p \) is assigned to location \( i \). Then \( \sum_{t=1}^{\bar{\tau}} S_{ip}^\tau \cdot \tau = \bar{\tau} \), and \( \sum_{t=1}^{\bar{\tau}} S_{ip}^\tau \cdot \tau \) extracts the start time step when resource \( p \) is assigned to location \( i \). Similarly, \( \sum_{t=1}^{T} E_{ip}^\tau \cdot \tau \) extracts the end time step (denoted as \( \bar{\tau} \)) of the assignment for resource \( p \) at location \( i \). The difference between these two values indicates the total number of time steps that resource \( p \) is assigned to location \( i \). Therefore, inequality (20) is equivalent to \( \bar{\tau} - \bar{\tau} \geq H \) under the case that resource \( p \) is assigned to location \( i \), and thus the requirement in Assumption 2 is satisfied.

Also, note that if \( \sum_{t=1}^{\bar{\tau}} S_{ip}^\tau = 0 \), then \( \sum_{t=1}^{\bar{\tau}} S_{ip}^\tau \cdot \tau = \sum_{t=1}^{\bar{\tau}} E_{ip}^\tau \cdot \tau = 0 \) due to (16). Thus, both sides of inequality (20) are zero. Under this condition, constraints (20) are redundant.

**Proposition 1.** Suppose resource \( p \) is assigned to location \( i \); if \( S_{ip}^\tau = 1 \) and \( E_{ip}^\tau = 1 \), then the start time step \( \bar{\tau} \) is earlier than the end time step \( \bar{\tau} \).

**Proof.** This proposition is valid as inequality (20) implies that \( \bar{\tau} - \bar{\tau} \geq H > 0 \). \( \square \)

### 3.3.3. Resource reallocation transition time constraints

The last set of constraints corresponds to Assumption 3 in Section 2.1, describing the transition time required to move a resource to another location upon reallocation. Suppose the latest time step that resource \( p \) stays at location \( i \) is \( \tau \), i.e., \( E_{ip}^\tau = 1 \). For any other location \( j \neq i \), resource \( p \) cannot arrive at intersection \( j \) before a time interval \( c_{ij} \) that is predetermined as the required reallocation transition time from location \( i \) to location \( j \). Thus, the reallocation transition time requirement in Assumption 3 can be mathematically formulated as:

\[ S_{jp}^{\tau+1} \leq 1 - E_{ip}^\tau \quad \forall j \neq i \in \hat{N}, p \in P, \tau \in T, t = 1,2,...,c_{ij}. \] (21)

Based on (21), if \( E_{ip}^\tau = 1 \) then the right-hand-side of inequality (21) enforces \( S_{jp}^{\tau+1} = 0 \) for all locations \( j \neq i \), during the period from \( \tau + 1 \) to \( \tau + c_{ij} \); if \( E_{ip}^\tau = 0 \), constraint (21) becomes redundant as \( S_{jp}^{\tau+1} \) is a binary variable.

The reallocation transition time constraints include a large number of inequalities in the DRA model, and hence increase the problem size significantly. Note that constraints (21) are binary constraints. If the inequalities are summed up over the candidate locations and time steps \( t = 1,2,...,c_{ij} \), a compact expression is obtained:

\[ \sum_{j=1}^{M} \sum_{\tau=1}^{c_{ij}} S_{jp}^{\tau+1} \leq C \cdot (1 - E_{ip}^\tau) \quad \forall i \in \hat{N}, p \in P, \tau \in T. \] (22)

where \( C \geq \max_{j} \sum_{\tau=1}^{c_{ij}} \) is a large number representing the upper bound of the summation of travel time intervals from one potential location to any other location. If \( E_{ip}^\tau = 0 \), the inequality \( \sum_{j=1}^{M} \sum_{\tau=1}^{c_{ij}} S_{jp}^{\tau+1} \leq C \) is satisfied automatically. On the contrary, if \( E_{ip}^\tau = 1 \), inequality \( \sum_{j=1}^{M} \sum_{\tau=1}^{c_{ij}} S_{jp}^{\tau+1} \leq 0 \) enforces \( S_{jp}^{\tau+1} = 0 \) for all \( 1 \leq t \leq c_{ij} \), \( j \neq i \), such that resource \( p \) cannot be assigned to any location \( j \in \hat{N} \) and \( j \neq i \) for time steps \( t = 1,2,...,c_{ij} \). Therefore,
inequality (22) is equivalent to inequality (21). Note that due to its summation form inequality (22) provides a tighter cut and significantly reduces the problem size.

3.4. Modeling impacts from deployed resources

The impacts from the deployed resources are modeled as link capacity increase for links connecting to the intersections with assigned resources. In particular, if any resource is assigned to intersection \( i \) at time step \( \tau \), then all links with tail node \( i \) have an additional capacity denoted by \( \Delta q_{ij} \). Therefore, for any link \((i,j)\), its link capacity can be represented by \( q_{ij} + \sum_{p=1}^{P} y_{ip}^\tau \cdot \Delta q_{ij} \). If binary variable \( y_{ip}^\tau = 1 \), then the link capacity at time step \( \tau \) has an increase in value as \( q_{ij} + \Delta q_{ij} \). The link capacity constraint (6) can be revised correspondingly as the following inequality to describe the impacts from the deployed moveable resources:

\[
x_{ij}^\tau \leq q_{ij} + \sum_{p=1}^{P} y_{ip}^\tau \cdot \Delta q_{ij}, \quad \forall (i,j) \in \mathcal{L}, \tau \in \mathcal{T}.
\]

3.5. Summary of the model and discussion

The dynamic resource allocation constraints presented in the previous section, as well as the mass balance constraints (2)-(5) and capacity constraints (23) on link flows, remain linear in the proposed formulation. This linearity enables the DRA problem to be formulated as a MILP. The mathematical formulation of the DRA problem is summarized as follows:

\[
\min z = \sum_{\tau=1}^{T} \sum_{(j,s) \in \mathcal{L}} \tau \cdot x_{js}^\tau
\]

subject to:

\[
\sum_{\tau=1}^{T} \sum_{(i,j) \in \mathcal{L}} x_{ij}^\tau = b_i \quad \forall i \in S^+
\]

\[
\sum_{\tau=1}^{T} \sum_{(j,s) \in \mathcal{L}} x_{js}^\tau = \sum_{i \in S^+} b_i
\]

\[
\sum_{\tau=1}^{T} \sum_{(i,j) \in \mathcal{L}} x_{ji}^{\tau-n} - \sum_{\tau=1}^{T} \sum_{(i,j) \in \mathcal{L}} x_{ji}^\tau \geq 0 \quad \forall i \in \mathcal{N} \setminus \{S^+ \cup s\}, \forall t \in \mathcal{T}
\]

\[
\sum_{\tau=1}^{T} \sum_{(i,j) \in \mathcal{L}} x_{ji}^{\tau-n} - \sum_{\tau=1}^{T} \sum_{(i,j) \in \mathcal{L}} x_{ji}^\tau = 0 \quad \forall i \in \mathcal{N} \setminus \{S^+ \cup s\}
\]

\[
x_{ij}^\tau \leq q_{ij} + \sum_{p=1}^{P} y_{ip}^\tau \cdot \Delta q_{ij}, \forall (i,j) \in \mathcal{L}, \tau \in \mathcal{T}
\]

\[
y_{ip}^\tau = \sum_{\tau=1}^{T} S_{ip}^{\tau} - \sum_{\tau=1}^{T} E_{ip}^{\tau} \quad \forall i \in \hat{\mathcal{N}}, p \in \mathcal{P}, \tau \in \mathcal{T}
\]

\[
\sum_{p=1}^{P} y_{ip}^\tau \leq 1 \quad \forall i \in \hat{\mathcal{N}}, \tau \in \mathcal{T}
\]

\[
\sum_{i=1}^{M} y_{ip}^\tau \leq 1 \quad \forall p \in \mathcal{P}, \tau \in \mathcal{T}
\]
\[
\sum_{t=1}^{T} \sum_{p=1}^{P} S_{ip}^{t} \leq 1 \quad \forall i \in \hat{N}
\]
\[
\sum_{t=1}^{T} \sum_{p=1}^{P} E_{ip}^{t} \leq 1 \quad \forall i \in \hat{N}
\]
\[
\sum_{t=1}^{T} S_{ip}^{t} = \sum_{t=1}^{T} E_{ip}^{t} \quad \forall i \in \hat{N}, p \in \mathcal{P}
\]
\[
\sum_{t=1}^{T} E_{ip}^{t} \cdot \tau - \sum_{t=1}^{T} S_{ip}^{t} \cdot \tau \geq H \cdot \sum_{t=1}^{T} S_{ip}^{t} \quad \forall i \in \hat{N}, p \in \mathcal{P}
\]
\[
\sum_{j \neq i}^{M} \sum_{t=1}^{\tau} S_{jp}^{t+1} \leq \mathcal{C}(1 - E_{ip}^{t}) \quad \forall i \in \hat{N}, p \in \mathcal{P}, \tau \in \mathcal{T}
\]
\[
y_{ip}^{t} \in \{0,1\} \quad \forall i \in \hat{N}, p \in \mathcal{P}, \tau \in \mathcal{T}
\]
\[
S_{ip}^{t} \in \{0,1\} \quad \forall i \in \hat{N}, p \in \mathcal{P}, \tau \in \mathcal{T}
\]
\[
E_{ip}^{t} \in \{0,1\} \quad \forall i \in \hat{N}, p \in \mathcal{P}, \tau \in \mathcal{T}
\]
\[
x_{ij}^{t} \geq 0, \forall (i,j) \in \mathcal{L}, \tau \in \mathcal{T}
\]

The proposed MILP model for the DRA problem is established on one set of real variables \(x_{ij}^{t}\) and three sets of binary variables \(y_{ip}^{t}\), \(S_{ip}^{t}\), and \(E_{ip}^{t}\), as discussed in Section 3.1. Compared with the MILP model proposed in He and Peeta (2014), this model adopts EAF to determine the dynamic traffic flows, and two additional sets of auxiliary variables \(S_{ip}^{t}\), and \(E_{ip}^{t}\) are introduced to the model. Note that if all the \(y_{ip}^{t}\) variables are known, then the binary variables \(S_{ip}^{t}\) and \(E_{ip}^{t}\) are determined automatically. Because of this, the MILP model (24)-(41) can be reformulated to omit all auxiliary variables \(S_{ip}^{t}\), and \(E_{ip}^{t}\). By doing so, the number of binary variables in the formulation is reduced by a factor of three. The resulting reformulation can be more compact in terms of problem size.

However, there are benefits to include auxiliary variables \(S_{ip}^{t}\), and \(E_{ip}^{t}\) in the proposed MILP model (24)-(41). The main advantage of the MILP model (24)-(41) is that using auxiliary variables \(S_{ip}^{t}\) and \(E_{ip}^{t}\) can generate strong valid equalities (30) which is beneficial to identify the optimal solution efficiently. Equalities (30) provide a high degree of correlation between the decision variables \(y_{ip}^{t}\) and auxiliary variables \(S_{ip}^{t}\), and \(E_{ip}^{t}\) as discussed in Section 3.1.

In operations research literature, introducing the start time and end time binary variables is a common approach to model scheduling problems. For example, Arroyo and Conejo (2000) proposed a MILP formulation for the unit commitment problem (Garver 1962) in the power industry, where generators’ start-up and shut-down statuses are used as auxiliary variables to model the on/off status of generators. Although Carrión and Arroyo (2006) reformulated the problem by eliminating the auxiliary binary variable to have a compact formulation, it has been shown by Ostrowski et al. (2012) that including the start-up and shut-down binary variables decreases its computational burden in comparison with the model using only commitment binary variables (i.e., without the start-up and shut-down ones). Their finding illustrates that adding binary variables in scheduling problems helps strengthen the linear relaxations and hence the bounds used in the branch-and-bound algorithm to solve the MILP.

Another advantage of keeping auxiliary variables \(S_{ip}^{t}\) and \(E_{ip}^{t}\) in the MILP model is that the integrality constraints (38) on decision variables \(y_{ip}^{t}\) can be relaxed as in the following proposition.

**Proposition 2.** In the proposed MILP model (24)-(41) for the DRA problem, the integrality constraints (38) on decision variables \(y_{ip}^{t}\) can be relaxed as:
\[
0 \leq y_{ip}^{t} \leq 1 \quad \forall i \in \hat{N}, p \in \mathcal{P}, \tau \in \mathcal{T}
\]

**Proof.** For any resource \(p\), at location \(i\) at time step \(\tau\), \(\sum_{t=1}^{T} S_{ip}^{t}\) can be either zero or one due to constraint (33), and \(\sum_{t=1}^{T} E_{ip}^{t}\) can be either zero or one as well due to constraint (34). In addition, inequality \(\sum_{t=1}^{T} E_{ip}^{t} \leq \sum_{t=1}^{T} S_{ip}^{t}\) is true.
due to Proposition 1. Therefore, \( y^r_{ip} = 0 \) if \( \sum_{t=1}^{T} y^r_{ip} = \sum_{t=1}^{T} S^r_{ip} \) and \( y^r_{ip} = 1 \) if \( \sum_{t=1}^{T} E^r_{ip} < \sum_{t=1}^{T} S^r_{ip} \), due to constraint (30). The integrality constraints (38) on \( y^r_{ip} \) can be relaxed by box constraints \( 0 \leq y^r_{ip} \leq 1 \).

\[ \text{4. Solution algorithm} \]

Discrete network design problems, including those for evacuation planning, are typically NP-hard. Heuristic methods are typically used as solution approaches to address such combinatorial optimization problems. For example, Tabu search is used to solve the contraflow problem (Tuydes and Ziliaskopoulos 2006; Xie and Turnquist 2011). Meanwhile, decomposition schemes have been applied to reduce problem complexity and develop efficient solution algorithms. For example, Li et al. (2003) applied the Dantzig-Wolfe decomposition scheme to develop an efficient algorithm to solve CTM-based SO-DTA.

Solving the DRA problem is more difficult than the classical static network design problem in evacuation planning as the time dimension is included. In He and Peeta (2014), an approximation approach was developed based on a two-stage optimization program formulation. However, it cannot guarantee that the solution will converge to optimality though it illustrates good computational efficiency.

Realizing the separable structure of the MILP model (24)-(41), we propose here to apply the Benders decomposition scheme to solve the DRA problem. The separability of the MILP model (24)-(41) can be analyzed as follows. In model (24)-(41), the evacuation link flow variables \( x^p \) are associated with a set of constraints (25)-(29) corresponding to the EAF problem. Constraints (30)-(37) are related to the more difficult decision variables \( y^r_{ip}, S^r_{ip}, \) and \( E^r_{ip} \) for determining the dynamic resource allocation. Only one set of constraints, i.e. capacity constraints (29), involve both \( x^p \) and \( y^r_{ip} \). Constraints (25)-(29) can be represented by a concise matrix form:

\[
\begin{bmatrix}
B & 0 \\
I & -Q
\end{bmatrix} \begin{bmatrix}
x \\ \gamma
\end{bmatrix} \leq \begin{bmatrix}
b \\ q
\end{bmatrix}
\]

where \( \gamma \) is a vector that includes \( y^r_{ip}, S^r_{ip}, \) and \( E^r_{ip} \), the first row of the block matrix describes constraints (25)-(28), and the second row of the block matrix represents the capacity constraints (29). Define matrix \( A = [B \quad I] \), \( F(\gamma) = [0 \quad -Q\gamma] \) and vector \( g = [b \quad q]^T \). The feasible set for binary variables \( y^r_{ip}, S^r_{ip}, \) and \( E^r_{ip} \), is defined as:

\[ \Gamma = \{ y^r_{ip}, S^r_{ip}, E^r_{ip} \mid \text{Constraints (30)-(40)} \} \] (44)

Then we have a concise representation of MILP model (24)-(41) corresponding to (7).

Note that for any given feasible dynamic resource allocation plan \( \gamma \in \Gamma \), solving the problem:

\[ \min_x \left\{ w^T x \mid Ax \leq g - F(\gamma), x \geq 0 \right\} \] (45)

is equivalent to solving the EAF problem (1)-(6) with the revised capacity specified by Eq. (29). By solving the dual problem of (45), the dual values to the constraints (29) can provide the information on the system travel time reduction if the capacity of link \((i,j)\) increases by one unit at time step \(\tau\). Denote \( u^*_{ij} \) as the dual variable for constraint (29). According to duality theory, if a specific constraint (29) is binding, then \( u^*_{ij} < 0 \). It implies that assigning a resource to the corresponding downstream intersection \(i\) at time step \(\tau\) will be beneficial to reduce the total system travel time. If constraint (29) is not binding, then \( u^*_{ij} = 0 \), meaning that assigning a resource would not be helpful to improve the evacuation network performance. Thus, the dual variables to constraints (29) provide the critical information to select the start time step and end time step of the dynamic resource allocations.

As discussed in Section 2.3, Benders decomposition requires solving the dual problem of (45), which is no longer the EAF problem. To maintain the EAF structure, which is easier to solve, we apply the dual simplex method to solve the EAF problem, and thus obtain the dual vector \( u \) associated with Eq. (29) instead of solving the dual problem directly. The set of dual vectors are used to compute \( (g - F(\gamma))u \) which further generates the Benders optimality cut. Note that we always generate a feasible dynamic resource allocation plan \( \gamma \in \Gamma \) in the primal problem. Thus the dual problem of (45) is always bounded. Hence, the auxiliary problem in the proposed Benders decomposition scheme reduces to using the dual simplex method to solve the EAF problem, which is a minimum cost network flow problem on the time-expanded network.

We can construct the solution algorithm for the MILP model (24)-(41) by following the Benders decomposition scheme presented in Section 2.3, as follows:
Step 0: Select $\varepsilon > 0$ for the stop criterion. Construct the restricted master problem (RMP) as follows: \( \min z \) subject to constraints (30)-(40). Solve the RMP, obtain an initial (feasible) dynamic resource allocation plan \( \gamma^0 \); set the iteration counter \( n = 0 \); initialize upper bound \( UB = +\infty \) and lower bound \( LB = -\infty \);

Step 1: Set counter \( n = n + 1 \). Update the link capacity based on the fixed dynamic resource allocation variables \( \gamma^r \). Use the dual simplex method to solve the EAF problem (24) - (29) and obtain dual vector \( u^r \) associated with equation (29). Compute \( (g - F(\gamma^r)) u^r \), and update \( UB = \min \{ UB, (g - F(\gamma^r)) u^r \} \).

Step 2: If \( (UB - LB)/UB < \varepsilon \), stop. The problem is solved to optimality. Otherwise, go to Step 3;

Step 3: Add the constraint \( z \geq (g - F(\gamma^r)) u^r \) to the RMP. Solve the updated RMP to determine the optimal solution \( \gamma^* \) and objective value \( z^* \). Let \( LB = z^* \). Go to Step 1.

In Step 1 of each iteration, solving the EAF problem identifies a dynamic traffic flow pattern with a reduced total system travel time based on the improved dynamic resource allocation plan. Therefore, the value of objective function of the EAF problem provides an improved upper bound to the evacuation. The elements in the dual vector \( u^r \) obtained by solving the EAF problem (24)-(29) indicate the benefits of assigning the resources to specific locations and time steps. This information is used to introduce the optimality cut \( z \geq (g - F(\gamma^r)) u^r \) into the RMP in Step 3. As a subset of constraints related to dynamic traffic flows are considered in solving the RMP, the value of objective function of the RMP provides a lower bound to the evacuation. The lower bound would likely increase when one more constraint is introduced to the RMP in each iteration. If the relative gap between the upper and lower bounds is less than a pre-specified threshold value, the optimal dynamic resource allocation plan is identified.

The main computational burden of this solution algorithm lies in Step 3, where a mixed integer linear program (MILP) needs to be solved. The introduction of the auxiliary variables \( S^r_{\mu} \) and \( E^\gamma_{\mu} \) substantially reduces the complexity of solving the MILP. By determining the starting and ending time index variables \( S^r_{\mu} \) and \( E^\gamma_{\mu} \), the dynamic resource allocation variables \( y^r_{\mu} \) are obtained automatically due to equation (30). This significantly reduces the computational complexity induced by the combinations of binary variables \( y^r_{\mu} \). The computational burden of solving the EAF problem in Step 1 is negligible as it has a classical network flow structure that can be solved efficiently by the network simplex method.

5. Numerical example

The Dallas-Fort Worth network, shown in Fig. 2, is used to perform the numerical analysis of the DRA model. The network consists of 180 nodes, 445 links and 13 traffic analytic zones (TAZ). The size of the network, in terms of total miles, is approximately 122 miles. Each time step consists of 30 seconds, and the total horizon is 100 minutes, or 200 time steps.

The network consists of a freeway, arterials and local streets. A freeway between nodes 116 and 117 is connected to arterials by a set of on- and off-ramps. The freeway has capacity 1,800 veh/hr/lane and free flow speed 65 miles/hr, and arterials have capacity 900 veh/hr/lane and free flow speed 40 miles/hr. On- and off-ramps have capacity 1200 veh/hr/lane and free flow speed of 30 miles/hr. With the allocation of resources at intersections on arterials and on- and off-ramps, the total capacity is 300 veh/hr/lane.

The total evacuation demand is 14,000, and shown in Table 1. All the boundary nodes, as highlighted in Fig. 2, are destinations. There are 5 resources (trained police personnel) who could be allocated at 30 locations which are highlighted Fig. 2. The DRA model is solved by CPLEX 12.5 in the Python interface. The relative gap of the stop criterion is 1%, i.e., \( (UB - LB)/UB < 1\% \). The computational environment consists of an Intel Xeon 16-Core E5260 Processor 2.4 GHz with 16GB of RAM.

The EAF problem is solved using the dual simplex method on the time-expanded network. Without resource allocation, the network takes 56 minutes to be cleared, and the total system travel time is 5,162.7 vehicle-hours. With resource allocation, the network takes 44 minutes to be cleared, and the total system travel time is 4,514.8 vehicle-hours. The improvement of clearance time is about 21%, and the improvement of total system travel time is about 12.5%. At convergence, the upper bound is 541,776, the lower bound is 540,485, and the relative gap is 0.2%. The arrival flows at the destination with and without resource allocation are plotted in Fig. 3. The majority of the computational time is used to solve the RMP, which is a mixed integer program. The time to solve the EAF is negligible. It generates three Benders optimality cuts to reach the relative gap of 0.2%, which indicates that the generated Benders cuts are effective.
Location can deploy dynamic resource

Destination

Fig. 2. Dallas-Fort Worth network.
Table 1. Evacuation demand.

| Node | Demand |
|------|--------|
| 114  | 2,000  |
| 88   | 2,000  |
| 146  | 3,000  |
| 85   | 3,000  |
| 94   | 4,000  |

The numerical example illustrates that the computational efficiency is significantly improved compared with the DRA model built upon the CTM. As reported in He and Peeta (2014), a DRA problem was considered on a grid network with 30 nodes and 107 links, and four moveable resources were assigned to 24 potential locations to assist the evacuation of 10,000 evacuees in one hour. In that example, CPLEX failed to identify a solution to satisfy the stopping criterion of 1% gap after more than 300 hours of computational time. In comparison, in this study, CPLEX took only 3.5 hours to identify the solution with 0.2% optimality gap, by applying the proposed MILP model and the Benders decomposition scheme.

The resource allocation is plotted in Fig. 4. The first resource is allocated to node 85 for the first 28 minutes. It then moves to node 152 for the period 30-36.5 minutes, and then moves to node 99 for the period 37-42.5 minutes. The second resource is allocated to node 88 for the first 31 minutes, and then moves to node 110 for the period 36-41 minutes. The third resource is allocated to node 114 for the first 20 minutes, and then moves to node 95 for the period 23-29 minutes, and then moves to node 149 for the period 30-37 minutes. The fourth resource is allocated to node 94 for the period 0-40 minutes, and the fifth resource is allocated to node 146 for the period 0-42 minutes. As the demand at nodes 94 and 146 is relatively high, the fourth and fifth resources are allocated there to facilitate evacuation without moves. The first three resources move once or twice. As the demand at nodes 88 and 114 is relatively low (2,000), the second and third resources are allocated to nodes 88 and 114 initially, and after those demands are cleared they are moved to nodes 110 and 149 to facilitate the demand at node 94 which is relatively high. Each resource is allocated at the same location for at least 5 minutes, as specified in the minimal time span constraint (36). The solution illustrates that the DRA problem can dynamically allocate resources to speed up the evacuation process on a large-scale network.
6. Concluding remarks

This study advances both methodological and computational research on the dynamic resource allocation problem. On the methodological side, this paper proposes a MILP formulation by applying the EAF to determine dynamic evacuation traffic flows. Additional auxiliary binary variables are introduced into the formulation to enhance the computational efficiency. These variables determine the start and end time steps of the schedule of a moveable resource assigned at a specific location, and assist in determining the presence of the resource through a set of strong valid equalities. With the introduced auxiliary variables, the binary variable $y_{i\tau}^r$ is relaxed to continuous variables. Further, as the start and end time steps of a resource are allocated at most once for the entire time horizon at any location, stronger cuts are obtained which reduce the feasible set substantially compared to the CTM-based model in He and Peeta (2014).

On the computational side, a solution algorithm is developed to seek optimal solutions efficiently to facilitate the application of the DRA model for large-scale transportation network evacuation planning. Compared with existing solution algorithms for solving DRA problem, the proposed solution algorithm can solve an optimal resource allocation plan for large-scale networks in reasonable time in a planning context. Further, it provides a benchmark to develop resource deployment strategies to support transportation network evacuation in the operational stage. The Benders decomposition scheme was implemented to decompose the MILP model into two sub-problems. The first sub-problem is the restricted master problem, which solves for the feasible resource allocation plan. The second sub-problem is the EAF problem with capacity improvement specified by resource allocation provided by the RMP. We use the dual simplex method to solve the EAF on the time-expanded network and the optimal dual variables are used to generate Benders optimality cuts. With the Benders decomposition scheme, the DRA problem can be solved for the optimal solution on large-scale networks.
The study offers the possibility of modeling and solving a range of evacuation problems to generate evacuation plans. First, the proposed dynamic resource allocation problem and its solution can be extended by integrating demand uncertainty into the modeling. The effectiveness of dynamic resource allocation depends on the evacuation demand distribution which is uncertain in real evacuation scenarios. Thus, modeling the demand uncertainty in the DRA problem can further improve the robustness of DRA plans. Second, the DRA can be further formulated for multi-stage evacuation planning, where the DRA plan is more sensitive to the temporal profile of the evacuation flow. Third, the proposed decomposition framework can potentially be applied to solve for the optimal solution for other bi-level programming problems in the context of evacuation planning and operations, for instance, network design, contraflow design, and combined signal and routing optimization, etc. Hence, the MILP model developed in this study can be treated as a general framework to solve other problems characterized by a similar bi-level structure with the system optimum objective.

Acknowledgements

This work is based on funding provided by the U.S. Department of Transportation through the NEXTRANS Center, the USDOT Region 5 University Transportation Center. The authors are solely responsible for the contents of this paper.

References

Anderson, D., 1972. Models for determining least-cost investments in electricity supply. The Bell Journal of Economics and Management Science 3, 267-299.
Arroyo, J.M., Conejo, A.J., 2000. Optimal response of a thermal unit to an electricity spot market. IEEE Transactions on Power Systems 15(3), 1098-1104.
Bakal, D. L., Smith, J.M., 1996. Resource allocation in state-dependent emergency evacuation networks. European Journal of Operational Research 89(3), 543-555.
Benders, J.F., 1962. Partitioning procedures for solving mixed variables programming problems, Numerische Mathematik 4, 238-252.
Bertsimas, D., Tsitsiklis, J.N., 1997. Introduction to Linear Optimization (Vol. 6). Belmont, MA, Athena Scientific.
Cai, X., McKinney, D.C., Lasdon, L.S., Watkins Jr, D.W., 2001. Solving large nonconvex water resources management models using generalized benders decomposition. Operations Research 49(2), 235-245.
Carrión, M., Arroyo, J.M., 2006. A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem. IEEE Transactions on Power Systems 21(3), 1371-1378.
Cerisola, S., Ramos, A., 2002. Benders decomposition for mixed-integer hydrothermal problems by Lagrangean Relaxation. In: Proceedings of 14th Power System Computation Conference, Sevilla, pp. 69-71.
Chen, M., Chen, L., Miller-Hooks, E., 2007. Traffic signal timing for urban evacuation. Journal of Urban Planning and Development 133(1), 30-42.
Chiu, Y. C., Zheng, H., Villalobos, J., Gautam, B., 2007. Modeling no-notice mass evacuation using a dynamic traffic flow optimization model. IIE Transactions 39(1), 83-94.
Cova, T.J., Johnson, J.P., 2003. A network flow model for lane-based evacuation routing. Transportation Research Part A 37(7), 579–604.
Daganzo, C., 1994. The cell transmission model. Part I: A simple dynamic representation of highway traffic. Transportation Research Part B 28(4), 269–287.
Daganzo, C., 1995. The cell transmission model, Part II: Network traffic. Transportation Research Part B 29(2), 79–93.
Dixit, V., Wolshon, B., 2014. Evacuation traffic dynamics. Transportation Research Part C 49, 114-125.
Fiedrich, F., Gebhauer, F., Rickers, U., 2000. Optimized resource allocation for emergency response after earthquake disasters. Safety Science 35(1), 41-57.
Fisher, M.L., Jaikumar, R., 1981. A generalized assignment heuristic for vehicle routing. Networks 11(2), 109-124.
Fleischer, L.K., 2001. Faster algorithms for the quickest transshipment problem. SIAM Journal on Optimization 12(1), 18-35.
Fleischer, L., Skutella, M., 2007. Quickest flows over time. SIAM Journal on Computing 36(6), 1600-1630.
Florian, M., Bushell, G., Ferland, J., Guerin, G., Nastansky, L., 1976. The engine scheduling problem in a railway network. INFOR 14(2), 121-138.
Gale, D., 1959. Transient flows in networks. Michigan Mathematical Journal 6(1), 59-63.
Garver, L.L., 1962. Power generation scheduling by integer programming—development of theory. Transactions of the American Institute of Electrical Engineers Part III: Power Apparatus and Systems 81(3), 730-734.
Geoffrion, A.M., 1972. Generalized benders decomposition. Journal of Optimization Theory and Applications 10(4), 237-260.
He, X., Peeta, S., 2014. Dynamic resource allocation problem for transportation network evacuation. Networks and Spatial Economics 14(3), 505-530.
Hsu, Y-T. Peeta, S., 2014a. Risk-based spatial zone determination problem for stage-based evacuation operations. Transportation Research Part C 41, 73-89.
Hsu, Y.-T. Peeta, S., 2014b. Behavior-consistent information-based network traffic control for evacuation operations. Transportation Research Part C 48, 339-359.

Jabari, S.E., He, X., Liu H.X., 2012. Heuristic solution techniques for no-notice emergency evacuation traffic management. In: Levinson, D.M., Liu, H.X., Bell, M. (Eds.), Network Reliability in Practice, pp. 241–259. Springer.

Jarvis, J.J., Ratliff, H.D., 1982. Note—some equivalent objectives for dynamic network flow problems. Management Science 28(1), 106-109.

Kalafatas, G., Peeta, S., 2009. Planning for evacuation: Insights from an efficient network design model. ASCE Journal of Infrastructure Systems 15(1), 21-30.

Kalafatas, G., Peeta, S., 2007. An exact graph structure for dynamic traffic assignment: Formulation, properties, and computational experience. In: Transportation Research Board 86th Annual Meeting (No. 07-1220).

Kongsomsaksakul, S, Yang, C, Chen, X., 2011. Shelter location-allocation model for flood evacuation planning. In: Proceedings of Transportation Research Board 93rd Annual Meeting (No. 14-3958).

Li, Y., Waller, S.T., Ziliaskopoulos, A., 2005. Shelter location-allocation model for flood evacuation planning. Journal of the Eastern Asia Society for Transportation Studies 6, 4237–4252.

Li, Y., Waller, S.T., Ziliaskopoulos, T., 2003. A decomposition scheme for system optimal dynamic traffic assignment models. Networks and Spatial Economics 3, 441–455.

Ng, M., Waller, S.T., 2010. A hybrid bilevel model for the optimal shelter assignment in emergency evacuations. Computer Letters 1(2), 111-119.

Ng, M., Park, J., Waller, S.T., 2010. A hybrid bilevel model for the optimal shelter assignment in emergency evacuations. Computer and Infrastructure Engineering 5(8), 547-556.

Peeta, S., Ukkusuri, S.V., 2007. System-optimal stochastic transportation network design. Transportation Research Record 2029, 80–86.

Peeta, S., Salman, F.S., Gunec, D., Viswanath, K., 2010. Pre-disaster investment decisions for strengthening a highway network. Computers & Operations Research 37(10), 1708–1719.

Sherali, H.D., Carter, T.B., Hobelka, A.G., 1991. A location-allocation model and algorithm for evacuation planning under hurricane/flood conditions. Transportation Research Part B 25(6), 439–452.

Sherali, H.D., Carter, T.B., Hobelka, A.G., 1991. A location-allocation model and algorithm for evacuation planning under hurricane/flood conditions. Transportation Research Part B 25(6), 439–452.

Thomas, P. R., Sahli, S., 1998. A Tabu search approach for the resource constrained project scheduling problem. Journal of Heuristics 4(2), 123-139.

Tuydes, H., Ziliaskopoulos, A., 2006. Tabu-based heuristic approach for optimization of network evacuation contraflow. Transportation Research Record 1964(1), 157-168.

Xie, C., Lin, D.Y., Waller, S.T., 2010. A dynamic evacuation network optimization problem with lane reversal and crossing elimination strategies. Transportation Research Part E 46(3), 295-316.

Xie, C., Turnquist, M.A., 2011. Lane-based evacuation network optimization: An integrated Lagrangian relaxation and Tabu search approach. Transportation Research Part C 19(11), 40–63.

Zheng, H., Chiu, Y.-C., Mirchandani, P., 2015. On the system optimum dynamic traffic assignment and earliest arrival flow problems. Transportation Science 49(1), 13-27.