Transportation Research Part B 71 (2015) 71–84

Contents lists available at ScienceDirect

Transportation Research Part B

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

Train design and routing optimization for evaluating criticality of freight railroad infrastructures

Abdullah A. Khaled [a], Mingzhou Jin [b,*], David B. Clarke [c], Mohammad A. Hoque [d]

[a] Mississippi State University, United States
[b] Department of Industrial and Systems Engineering, The University of Tennessee, Knoxville, 525D John D. Tickle Engineering Building, 851 Neyland Drive, Knoxville, TN 37996, United States
[c] The University of Tennessee, Knoxville, United States
[d] Texas Southern University, United States

ARTICLE INFO

Article history:
Received 2 June 2013
Received in revised form 30 September 2014
Accepted 3 October 2014
Available online 15 November 2014

Keywords:
Railroads
Freight transportation
Security and emergencies
Optimization

ABSTRACT

Freight transportation by railroads is an integral part of the U.S. economy. Identifying critical rail infrastructures can help stakeholders prioritize protection initiatives or add necessary redundancy to maximize rail network resiliency. The criticality of an infrastructure element, link or yard, is based on the increased cost (delay) incurred when that element is disrupted. An event of disruption can cause heavy congestion so that the capacity at links and yards should be considered when freight is re-routed. This paper proposes an optimization model for making-up and routing of trains in a disruptive situation to minimize the system-wide total cost, including classification time at yards and travel time along links. Train design optimization seeks to determine the optimal number of trains, their routes, and associated blocks, subject to various capacity and operational constraints at rail links and yards. An iterative heuristic algorithm is proposed to attack the computational burden for real-world networks. The solution algorithm considers the impact of volume on travel time in a congested or near-congested network. The proposed heuristics provide quality solutions with high speed, demonstrated by numerical experiments for small instances. A case study is conducted for the network of a major U.S. Class-I railroad based on publicly available data. The paper provides maps showing the criticality of infrastructure in the study area from the viewpoint of strategic planning.

1. Introduction and literature review

Railroads play a significant role in U.S. freight transportation to support the nation's supply chain and national security (Davis et al., 2012; Stodolsky, 2011). Total U.S. rail freight ton-miles have doubled and density (measured by total ton-miles per mile of track) tripled between1980 and 2006 (Cambridge Systematics, 2007). During the same period, total ton-miles carried by Class-I railroads increased by 93 percent (Eakin et al., 2009). In contrast, freight rail infrastructure has shrunk as carriers sought to reduce costs. Rail miles have decreased by 42 percent since 1980. In consequence, the U.S. freight railroad industry is currently operating without much excess capacity. A study by the American Association of State Highway and Transportation Officials (AASHTO) shows that projected rail freight movement will increase by 84 percent by 2035 (AASHTO, 2007). To promote railway freight security and resilience, the Transportation Security Administration (TSA) has

* Corresponding author. Tel.: +1 (865) 974 9992; fax: +1 (865) 974 0588.
E-mail address: jin@utk.edu (M. Jin).

http://dx.doi.org/10.1016/j.trb.2014.10.002
0191-2615/

established a freight rail division with the vision of "ensuring the secure movement of all cargo on our nation's freight rail systems and promoting the free flow of commerce by working with our public and private sector partners to maintain a secure, resilient, and sustainable network" (TSA, 2012). To support this mission, rail infrastructure of high consequence and vulnerability need to be identified so that rail stakeholders can take effective measures to protect critical infrastructure.

This paper proposes a solution approach to determine the criticality of railroad network infrastructure at the strategic level. The criticality of an infrastructure element, a node or a link, is evaluated by estimating the increased cost (delay) when this element is disrupted. After a disruption, traffic needs to be re-routed over the residual network to keep freight flow. The disruption may cause huge congestions, directly or indirectly, over a big part of the network and make the flows at many links and yards close to their capacities. A system-wide optimization model and a solution approach considering the congestion effects and capacity restrictions are therefore necessary to route traffic and further evaluate the effect of a disruption.

Railroad operations are highly complex so that real-world problems are large in size (Assad, 1981). Early optimization attempts had limited success and could not incorporate many of real-life characteristics (Assad, 1980b; Haghani, 1987; Cordeau et al., 1998; Newman et al., 2002). In recent years, railroads have started to implement optimization-based decision support systems to address various operational issues, such as blocking planning, Block-to-Train Assignment (BTA), train scheduling, locomotive scheduling, crew scheduling, empty car movement, and so on. Modern optimization models and solution approaches are founded on more sophisticated problem formulations, the ever-increasing computing capacity, and much improved data (Brannlund et al., 1998; Barnhart et al., 2000; Ahuja et al., 2007; Jha et al., 2008). Freight rail operations often start by developing a blocking plan. It aggregates a vast number of shipments into *blocks* as they move together to reduce reclassification at yards and overall intermediate handling costs (Newton et al., 1998; Barnhart et al., 2000; Jha et al., 2008; Yue et al., 2011). Based upon a blocking plan, *train routing* (TR) or scheduling identifies train routes, timetables, and frequencies (Farvolden and Powell, 1994; Brannlund et al., 1998). BTA is then conducted to determine which trains should carry which blocks (Kwon et al., 1998; Jha et al., 2008). The combination of the two tasks of BTA and TR is called *train design*, which is a highly combinatorial and complex optimization problem. A number of efforts have been made to solve this problem using different cost terms and business constraints. Most of these researches divided the overall problem into two sub-problems of train design and block routing and solved them iteratively (Assad, 1980a; Haghani, 1987, 1989; Keaton, 1989, 1992; Marín and Salmerón, 1996). Gorman (1998) and Newman and Yano (2001) used the integrated approach to solve the whole problem. More recently, a few works addressed the large-scale train design problems (Ahuja et al., 2005; Lozano et al., 2011; Colombo et al., 2011; Wang et al., 2011; Jin et al., 2013). However, all above models or solutions approaches in literature studied train design under normal conditions without considering the capacities at yards and links and the congestion factor that may occur during any catastrophes. Cacchiani et al. (2014) recently provided an overview of real-time models and algorithms for recovery from disruptions or disturbances. In practice, North American railroads widely use the MultiModal system to create the sequence of blocks and train routes with both optimization and simulation techniques (Ireland et al., 2004). MultiModal creates block sequences by minimizing the number of switches and total car-miles under various constraints through shortest-path algorithms. The system is used for railway planning under normal conditions without considering congestion. In addition to freight train design, a lot of studies have been done for passenger train line planning and routing (e.g., Carey, 1994; Kaspi and Raviv, 2013), which do not have the classification issue but focus more on timetables. Caprara et al. (2011) provided an overview of optimization problems in passenger train planning.

Our research focuses on the train design problem in re-routing freight traffic to determine rail infrastructure criticality for the strategic planning purpose. The re-routing under disruptions is different from normal train design that disruptions could cause serious congestions at railroad links and nodes. This paper develops a train design approach that incorporates capacity constraints at links and yards, operational constraints of trains, and the volume-speed relationships at links. The criticality of one link or node is evaluated by the increased transportation cost (delay) caused by its disruption to the baseline case. Each disruption scenario is solved by the proposed train design approach.

The remainder of the paper is organized as follows. Section 2 presents an optimization model for train design and routing under congestion. Section 3 describes a solution approach based on a decomposition of the overall problem into sub-problems of BTA and TR and shows its performance with a small instance. A case study is performed for a real-world case on a major Class-I railroad in Section 4. Section 5 presents conclusion and discusses possible model extensions.

2. Mathematical formulation considering congestions

Different from train forming and routing under normal conditions, a disruptive event can cause a rail network to operate under high congestion and close to its capacity at both links and yards. An optimization model is necessary to incorporate the following congestion concerns. In the literature, the train design optimization models for a rail network under normal conditions do not consider the capacity at yards and links or speed-volume relationship at links (e.g., Colombo et al., 2011; Jin et al., 2013).

At each link, the travel time depends on the number of trains traveling on the link. When more trains are assigned to it, the speed on the link may be reduced. The decreasing speed-volume relationship depends on link features, such as the number of tracks, siding, and its signal system. Furthermore, the model also has an upper bound on the daily number of trains that can travel on each link. A recent research by Norfolk Southern (NS) Railroad (Yoon et al., 2011) tried to use a numerical way to calculate this daily upper bound on their rail links. Shih et al. (2014) studied capacity expansion on links by sparse sidings.

At each classification yard, increased volume may increase the waiting time for an inbound train to be humped but may decrease the waiting time to form an departing train so that the cost (or time) for a railcar to be classified at one specific yard is assumed to be independent from the volume. However, a yard can only handle a limited number of trains and/or cars. If the volume exceeds the capacity, the traffic may overflow and block the traffic on the main tracks. This assumption is consistent to the numerical results presented by Petersen (1977) and Turnquist and Daskin (1982) and simulation results by Marinov and Viegas (2009). Their results showed that the dwell time at a yard keeps almost the same until the volume reaches a threshold value (capacity). Beyond that value, the dwell time goes up rapidly. The capacity concern is not important under the current normal practice because freight trains in the U.S. usually follow a timetable, which does not change daily. Furthermore, most yards are currently operated within their capacities.[1]

The following train forming and routing model with congestion concerns considers a directed railway network $G = (N,A)$, where N is the set of nodes and $A = \{(i,j):i,j \in N\}$ is the set of directed links. $(i,j) \in A$ implies $(j,i) \in A$. N_c is the set of nodes for classification and $N_c \subseteq N$. At each classification node (also known as yard) $i \in N_c$, a capacity is defined by the number of trains and the number of cars that the yard can handle every day, U_i^T and U_i^C respectively. The cost (or dwell time) for a railcar at yard $i \in N_c$ is assumed at C_i^C, which is independent from the classification volume. The dwell time in a yard depends on the layout and connection plan of the yard, as shown by Petersen (1977). In practice, railroads have historical data and also often use simulation to estimate the capacity for a specific yard. The distance of each link $(i,j) \in A$ is d_{ij}. The travel cost (related to travel time) per railcar depends on the daily number of trains traveling on the link. If l trains travel on link (i,j) per day, including both directions, the travel cost per railcar is C_{ij}^l because more assigned trains usually mean lower speeds (Lai and Barkan, 2009), especially for single tracks. At most U_{ij}^M trains, on both directions, can be assigned to link (i,j), where $i < j$. Furthermore, the train length and weight are limited by U_{ij}^L and U_{ij}^W on each link $(i,j) \in A$, respectively, which are usually decided by link features. In addition, each train incurs a start cost of C^S and travel cost C_{ij}^T for passing link (i,j). C_{ij}^T is the product of the train travel cost C^T per mile and the link distance d_{ij}. Both C^S and C_{ij}^T are independent from the travel time and are therefore independent from the volume over links. All the four cost items $\left(C_i^C, C_{ij}^l, C_{ij}^T, C^S\right)$ are normalized into a momentary unit though several of them are originally based on ton-hours.

A set of blocks B need to be shipped from their origins to destinations, where b is its index. For each block b, we know the following information: number of railcars r_b, origin and destination o_b and d_b, length L_b, and weight w_b. Whenever a train stops to either pick up or drop off blocks, it is called a work event. Each train cannot exceed its maximum allowable intermediate work events R^{we} in its route, excluding the train origin and destination. Each train is restricted to carry a maximum number of blocks R^{bt} along its whole route. Without the loss of generality, this paper further makes the following assumptions.

- Each train can depart from a node at most once;
- A train can start and end at the same node; and
- A train can only pick up and drop off one block by at most once.

A mixed integer program (1)–(29) is developed to represent the described problem with the following decision variable definitions.

z_{ij}^t	whether train t travels along link $(i,j) \in A$;
y_{ij}^l	whether there are l trains traveling along link (i,j) on both directions, $i < j, l \in \{0, \ldots, U_{ij}^M\}$;
x_i^t	whether train t has a working event at node $i \in N_c$;
$u_i^t\,(v_i^t)$	whether train t starts (ends) at node $i \in N$;
a_i^t	whether train t starts <u>and</u> ends at node $i \in N$;
$p_i^{b,t}\left(q_i^{b,t}\right)$	whether train t picks up (drops) block b at node $i \in N_c$;
$g_i^t\left(h_i^t\right)$	length (tonnage) of train t after visiting node $i \in N$;
θ_i^t	number of railcars carried by train t after passing node $i \in N$;
γ_i^t	number of railcars carried by train t after being classified at yard $i \in N_c$;
φ_{ij}^t	number of railcars that train t carries along link $(i,j) \in A$;
τ_{ij}	total car travel cost over link $(i,j) \in A, i < j$, for both directions; and
k_i^t	an artificial variable for eliminating sub-tours, $i \in N, t \in T$.

[1] This yard capacity assumption was made based on authors' interactions with CSX and NS railroads, two U.S. Class-I railroads.

$$\text{Min} \sum_{t\in T}\sum_{(i,j)\in A} C_{ij}^T z_{ij}^t + C^S \sum_{t\in T}\sum_{i\in N} u_i^t + \sum_{i\in N_c}\sum_{t\in T} C_i^C \gamma_i^t + \sum_{(i,j)\in A, i<j} \tau_{ij} \tag{1}$$

$$\text{S.T.} \sum_{\forall j:(i,j)\in A} z_{ij}^t - \sum_{\forall j:(j,i)\in A} z_{ji}^t = u_i^t - v_i^t \quad \forall t\in T,\ \forall i\in N \tag{2}$$

$$\sum_{\forall j:(i,j)\in A} z_{ij}^t \leqslant 1 \quad \forall t\in T,\ \forall i\in N \tag{3}$$

$$\sum_{\forall j:(i,j)\in A} z_{ji}^t \leqslant 1 \quad \forall t\in T,\ \forall i\in N \tag{4}$$

$$\sum_{t\in T} z_{ij}^t + \sum_{t\in T} z_{ji}^t = \sum_{l\in\{0,\ldots,U_{ij}^M\}} l y_{ij}^l \quad \forall (i,j)\in A,\ i<j \tag{5}$$

$$\sum_{l\in\{0,\ldots,U_{ij}^M\}} y_{ij}^l = 1 \quad \forall (i,j)\in A,\ i<j \tag{6}$$

$$\sum_{i\in N} u_i^t \leqslant 1 \quad \forall t\in T \tag{7}$$

$$2a_i^t \leqslant v_i^t + u_i^t \quad \forall t\in T,\ \forall i\in N \tag{8}$$

$$k_j^t \geqslant k_i^t + 1 - Ma_j^t - \left(1 - z_{ij}^t\right)M \quad \forall t\in T,\ \forall (i,j)\in A \tag{9}$$

$$\begin{cases} \sum_{t\in T} p_i^{b,t} = 1 - \sum_{t\in T} q_i^{b,t} = 1 & \text{if } i = o_b \\ \sum_{t\in T} q_i^{b,t} = 1 - \sum_{b\in B} p_i^{b,t} = 1 & \text{if } i = d_b \\ \sum_{t\in T} p_i^{b,t} = \sum_{t\in T} q_i^{b,t} & \text{if } i\in N_c - o_b - d_b \end{cases} \quad \forall b\in B,\ \forall i\in N_c \tag{10}$$

$$u_i^t + x_i^t \geqslant p_i^{b,t} \quad \forall t\in T,\ \forall i\in N_c, \forall b\in B \tag{11}$$

$$v_i^t + x_i^t \geqslant q_i^{b,t} \quad \forall t\in T,\ \forall i\in N_c, \forall b\in B \tag{12}$$

$$u_i^t + x_i^t \leqslant \sum_{\forall j:(i,j)\in A} z_{ij}^t \quad \forall t\in T,\ \forall i\in N_c \tag{13}$$

$$\sum_{i\in N_c} p_i^{b,t} = \sum_{i\in N_c} q_i^{b,t} \quad \forall b\in B,\ \forall t\in T \tag{14}$$

$$k_j^t \geqslant k_i^t + 1 + M\left(p_i^{b,t} + q_j^{b,t} - v_j^t - 2\right) \quad \forall t\in T, \forall i, j\in N,\ \forall b\in B \tag{15}$$

$$g_j^t \geqslant g_i^t + \sum_{b\in B} p_j^{b,t} L_b - \left(1 - z_{ij}^t\right)M - \sum_{b\in B} q_j^{b,t} L_b \quad \forall t\in T,\ \forall (i,j)\in A \tag{16}$$

$$(u_i^t - 1)M + \sum_{b\in B} p_i^{b,t} L_b \leqslant g_i^t \leqslant \sum_{\forall j:(i,j)\in A} U_{ij}^L z_{ij}^t \quad \forall t\in T,\ \forall i\in N_c \tag{17}$$

$$h_j^t \geqslant h_i^t + \sum_{b\in B} p_j^{b,t} w_b - \left(1 - z_{ij}^t\right)M - \sum_{b\in B} q_j^{b,t} w_b \quad \forall t\in T,\ \forall (i,j)\in A \tag{18}$$

$$(u_i^t - 1)M + \sum_{b\in B} p_i^{b,t} w_b \leqslant h_i^t \leqslant \sum_{\forall j:(i,j)\in A} U_{ij}^W z_{ij}^t \quad \forall t\in T,\ \forall i\in N_c \tag{19}$$

$$\theta_j^t \geqslant \theta_i^t + \sum_{b\in B} p_j^{b,t} r_b - \left(1 - z_{ij}^t\right)M - \sum_{b\in B} q_j^{b,t} r_b \quad \forall t\in T,\ \forall (i,j)\in A \tag{20}$$

$$\theta_i^t \geqslant (u_i^t - 1)M + \sum_{b\in B} p_i^{b,t} r_b \quad \forall t\in T,\ \forall i\in N_c \tag{21}$$

$$\varphi_{ij}^t \geqslant \theta_i^t - M(1 - z_{ij}^t) \quad \forall t\in T,\ \forall (i,j)\in A \tag{22}$$

$$\sum_{t\in T} x_i^t + \sum_{t\in T} u_i^t \leqslant U_i^T \quad \forall i\in N_c \tag{23}$$

$$\gamma_i^t \geqslant \theta_i^t - M(1 - x_i^t - u_i^t) \quad \forall t\in T,\ \forall i\in N_c \tag{24}$$

$$\sum_{t\in T} \gamma_i^t \leqslant U_i^C \quad \forall i\in N_c \tag{25}$$

$$\sum_{i\in N_c} x_i^t \leqslant R^{we} \quad \forall t\in T \tag{26}$$

$$\sum_{b\in B, i\in N_c} p_i^{b,t} \leqslant R^{bt} \quad \forall t\in T \tag{27}$$

$$\tau_{ij} \geqslant C_{ij}^l \sum_{t\in T}\left(\varphi_{ij}^t + \varphi_{ji}^t\right) - M\left(1 - y_{ij}^l\right) \quad \forall (i,j)\in A,\ i<j,\ l\in\left\{0,\ldots,U_{ij}^M\right\} \tag{28}$$

$$\sum_{i\in N} i u_i^t \leqslant \sum_{i\in N} i u_i^{t+1} \quad t = 1,\ldots,T-1 \tag{29}$$

$$x_i^t, z_{ij}^t, u_i^t, v_i^t, p_i^{b,t}, q_i^{b,t}, a_i^t, y_{ij}^l \in \{0,1\}; g_i^t, h_i^t, \theta_i^t, \gamma_i^t, \varphi_{ij}^t, k_i^t, \tau_{ij} \geqslant 0$$

The objective function (1) minimizes the total cost, including train travel cost, train start cost, railcar classification costs, and railcar travel cost along links, which depends on traffic volume. Constraint set (2) keeps the train flow conservation at all nodes. Constraint sets (3) and (4) make sure that one train can leave each node (link) at most once. Constraint sets (5) and (6) are used to obtain the number of trains traveling over link (i,j), represented by y^l_{ij}. They also make sure that the number of trains on both directions over link (i,j) will not exceed its capacity of U^M_{ij}. Constraint set (7) guarantees that each used train only starts once while constraint set (8) permits a train to start and end at the same node. Sub-tours are eliminated by constraint set (9). When node j is visited by train t immediately after node i, k^t_j will be greater than k^t_i by at least 1. Here, M is a big number. Constraint set (10) guarantees a block is picked up (dropped off) by a train at either its origin (destination) or in the classification yard where some train drops (picks) it and a block is not dropped (picked) at its origin (destination). Constraint sets (11) and (12) together make $x^t_i = 1$ when train t picks or drops (except its terminating yard) some blocks at yard i. Constraint set (13) guarantees that a train definitely passes the node where the train has a work event (classification). When a train picks up a block, the train must drop it, as indicated by (14). A train must pick up a block before it can drop that block, guaranteed by constraint set (15) using the ordering variable of k^t_i obtained from constraint set (9). Please note that (15) does not matter when the dropping point is the termination node of the train, which is necessary when a train starts and terminates at the same node. The length of train t after visiting a node is obtained by constraint sets (16) and (17) and satisfies the maximum train length restriction on each link. Similarly, the weight restriction of trains over each link is met through constraint sets (18) and (19). The number of railcars a train has after visiting a node is obtained by constraint sets (20) and (21). The number of railcars that a train carries over a link is obtained by constraint set (22). Constraint set (23) restricts the number of trains that can be classified at each yard i while constraint sets (24) and (25) restrict the number of railcars each classification yard can handle. Constraint set (26) restricts the maximum number of intermediate work events a train can have. The number of blocks one train can carry cannot exceed the maximum number of blocks that a train is allowed to transport, restricted by constraint set (27). Constraint set (28) determines the railcar travel cost traveling both directions over each link, which is used in the objective to calculate the total railcar travel cost for all links. Finally, Constraint set (29) is used to reduce the symmetry of solutions to improve the computational efficiency.

In summary, the model of (1)–(29) has three major parts. Constraint sets (2)–(9) are for train routing. Constraint sets (10)–(15) are used to make sure all blocks are served well by trains. Constraint sets (16)–(27) are used to guarantee that various capacity restrictions are met, including yard capacity, link capacity, and the restriction of train length and weight on each link. This MIP model is a highly combinatorial optimization problem that cannot be solved for a large network using exact optimization techniques, though small instances can be solved optimally using commercial software (e.g., IBM CPLEX), as shown in Table 1 in Subsection 3.4. Hence, a heuristic algorithm is proposed in the next section to obtain quality solutions within a reasonable amount of computational time.

3. Solution approach

An iterative heuristic algorithm, illustrated by Fig. 1, is developed with two major sub-problems of BTA and TR. The inputs include network data, block data, and all the cost parameters and restrictions. The initialization stage creates a sub-network for each block and assigns the blocks with reasonably large number of railcars to 'unit' trains, which travel from their origins to destinations without any classifications. The BTA algorithm assigns the remaining blocks one by one to trains and provides a list of trains that are formed to serve all the blocks. At the TR stage, formed trains are routed optimally with their known stops at classification yards. BTA and TR stages are repeated until the percent cost gap between two consecutive TR runs falls below a predefined tolerance limit Δ.

Table 1
Results of the optimal and heuristic solution approaches for small instances.

Instance	Number of Nodes-Arcs-Blocks	CPLEX		Heuristic Solution		Difference	
		Value ($)	CPU Time (sec.)	Value ($)	CPU Time (sec.)	Value	CPU Time
1	4-5-4	15387.5	131.8	15387.5	0.11	0.000%	99.947%
2	5-7-5	17377.5	921.6	17377.5	0.19	0.000%	99.983%
3	6-9-6	23637.5	1838.2	23637.5	0.28	0.000%	99.987%
4	7-11-7	29328.8	4234.5	29328.8	0.39	0.000%	99.989%
5	8-13-8	40112.5	6958.9	40112.5	0.46	0.000%	99.992%
6	9-15-9	44098.2	9989.1	44098.2	0.59	0.000%	99.993%
7	10-17-10	47216.6	12874.0	47216.6	0.73	0.000%	99.994%
8	11-19-11	51462.3	18985.4	51462.3	0.91	0.000%	99.995%
9	12-21-12	54539.6[a]	20000.0	54445.8	1.01	0.172%	99.994%
10	13-23-13	56768.8[a]	20000.0	56135.5	1.28	1.115%	99.993%

[a] Best solution when the time limit of 20,000 s is reached.

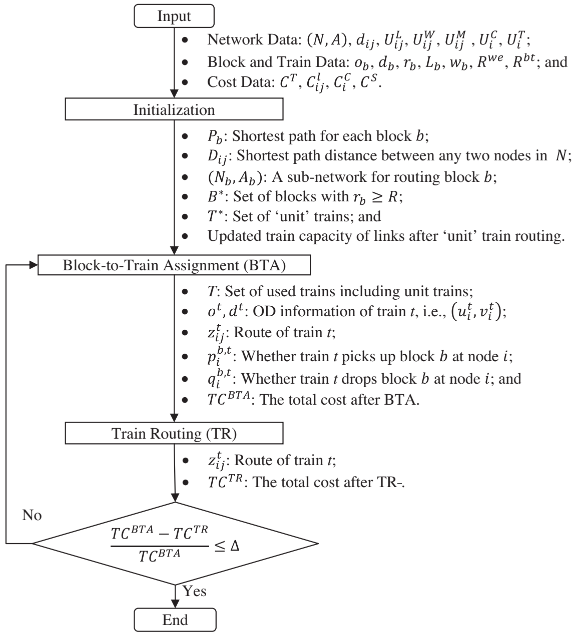

Fig. 1. Framework of the proposed heuristic algorithm.

3.1. Initialization Step

In the Initialization step, detailed in Fig. 2, a sub-network for each individual block from the whole network is built to reduce computational burden. Rather than the whole network (N,A), only a sub-network of (N_b,A_b) in which all nodes have a distance of less than D miles from the shortest path for the block are considered when routing block b. Fig. 3 shows an example of a 10-node network with links and their distances. If we consider a block b whose $o_b = 2$ to $d_b = 4$ (i.e. the block is from yard 2 to yard 4), the shortest path for block b is obviously $P_b = 2 \rightarrow 1 \rightarrow 4$ with the distance of 80. If we assume

***Initialization*:**

Create a shortest path for each pair of two nodes $i, j \in N$ and have the shortest distane D_{ij},

For each block b in B:

  Include all nodes in the shortest path P_b from o_b to d_b into set N_b;

  Add all nodes in $N \backslash N_b$ whose distance from any node in P_b is less than or equal to a given real number D into N_b (in the case, $D = 100$ miles); and

  Let A_b the set of links with both ends in N_b,

For each block b in with $r_b \geq R$ (in the case study, $R = 80$ railcars):

  Assign 'unit' trains for b, each following the shortest path P_b that satifies link weight and length restriction along the path;

  Update the avaialble train capacity U_{ij}^M at links $(i, j) \in P_b$ and U_i^T and U_i^C at the origin; and

  Update r_b, L_b, and w_b by deducting railcars carried by 'unit' trains.

Fig. 2. Steps of initialization algorithm for creating block sub-networks and 'Unit' trains.

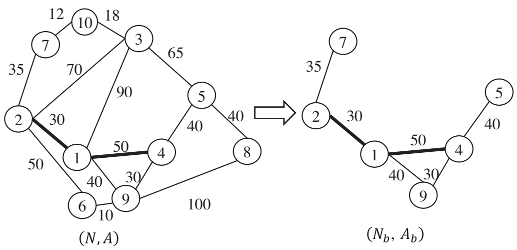

Fig. 3. Reduced network for block b with $o_b = 2$ and $d_b = 4$.

$D = 40$, we have $N_b = \{1,2,4,5,7,9\}$ and $A_b = \{(1,2),(2,1),(1,4),(4,1),(1,9),(9,1),(2,7),(7,2),(4,5),(5,4),(4,9),(9,4)\}$. The new reduced network (N_b,A_b) will be used later for routing block b. If the demand from node 2 to node 4 is 100 railcars, a unit train with $R = 80$ railcars is travelling directly from node 2 to node 4 without classification when the train does not violate the weight and length restrictions on links (2,1) and (1,4). The remaining 20 railcars will be considered as the demand for this OD pair. At the same time, the capacity at involved yards and links (i.e., (1,2) and (1,4)) are reduced.

3.2. Block-to-Train Assignment (BTA)

A set of trains T serving blocks $b \in B$ is assumed to be obtained from the previous iteration. The given information at the beginning of BTA is as follows.

For each train $t \in T$:

- The route from node o^t to node d^t, the set of visited nodes N^t, and the set of visited links A^t,
- The number of carried railcars θ_i^t, length g_i^t, and weight h_i^t after visiting node $i \in N^t$, and
- The number of work events X^t, the set of carried blocks B^t, and their pickup and drop-off points.

For each classification yard $i \in N_c$:

- The number of classified trains, λ_i, and
- The number of classified railcars, γ_i.

For each link $(i,j) \in A$:

- The total number of railcars carried by all trains, φ_{ij}, and
- The total number of traveled trains, l_{ij}.

BTA, explained in Fig. 4, reassigns blocks in B, one by one following a random order, to existing trains or a new train with the least cost. The random order in Step 1 is used to avoid local optimal solutions over iterations. Train information is updated along with the network capacity consumptions after assigning a block. At Step 3, the set of candidate trains T^b that can carry block $b \in B$ is formed by deleting the trains that (1) have carried R^{bt} blocks belonging to $B - \{b\}$, (2) have carried only block b, (3) are unit trains, or satisfy (4) $A^t \cap A_b = \varnothing$ (called the train selection criteria in Fig. 4). Here, A_b is the candidate links for block b and obtained in the *Initialization* step.

A new variable of ρ_i^t is introduced to indicate whether train t is classified at node $i \in N^t$ in Step 4.

$$\rho_i^t = \begin{cases} 0 & \text{if train } t \text{ pickups/drops any blocks in } B - \{b\} \text{ at node } i \in N^t \\ 1 & \text{otherwise.} \end{cases}$$

At step 4, a new directed network for each train $t \in T^b$ is built for block b that is under consideration. If $o_b \in N^t$ and/or $d_b \in N^t$, we delete all nodes visited before o_b and all nodes visited after d_b from N^t to have N_b^t and delete all links visited before o_b or after d_b from A^t to have A_b^t. If $A_b^t = \varnothing$, we don't build a network and skip all future steps for the train. We then delete any links that will exceed its length and weight maximum if train t carries block b from A_b^t. For each pair of nodes $m,n \in N_b^t$, when there is a path from m to n in (N_b^t, A_b^t) and $\rho_m^t + \rho_n^t + X^t \leqslant R^{we}$, $\lambda_m + \rho_m^t \leqslant U_m^T$, $\gamma_m + \rho_m^t\theta_m^t \leqslant U_m^C$ and $\gamma_n + \rho_n^t\theta_n^t \leqslant U_n^C$, a new link directly from m to n called (m,n) is added into A_b^t with the cost CN_{mn}^t based on the shortest path from m to n in the network of

***BTA*:**

Step 1: Randomize the blocks in B.

Step 2: Select the next block $b \in B$ following the above order.

Step 3: Create a set of candidate trains T^b following the train selection criteria.

Step 4: For each train $t \in T^b$

- Build the sets of nodes N_b^t and links A_b^t that can be used for block b and train t.
- Calculate the cost of using train t to carry block b from node i to node j, both belonging to N_b^t, CN_{ij}^t.
- Extend the network of (N_b^t, A_b^t) to o_b (or d_b) if o_b (or d_b) in not in N_b^t and $d^t \neq o^t$ and calculate the cost of using train t to pick up b from o_b (or drop b at d_b) by extending the train route.

Step 5: Add a new route (train) from o_b to d_b to handle block b.

Step 6: Build a network by combining all (N_b^t, A_b^t), $\forall t \in T^b$.

Step 7: Find the shortest cost path from o_b to d_b in the combined network.

Step 8: Update the train information and available network capacity after assigning block b.

Step 9: if there are still blocks have not be considered, go to Step 2; otherwise, STOP.

Fig. 4. Block-to-Train Assignment (BTA) Algorithm.

(N_b^t, A_b^t). Adding this link indicates that train t to carry block b from m to n without violating any constraints. The total cost of using train t to carry block b from m to n is denoted by CN_{mn}^t and can be calculated by (30).

$$CN_{mn}^t = DC_{mn}^t r_b + (\rho_m^t \theta_m^t + r_b)C_m^C + \rho_n^t \theta_n^t C_n^C + \sum_{k \in P_{m\to n}-\{m\}-\{n\}} (1-\rho_k^t) r_b C_k^C. \tag{30}$$

The shortest path cost DC_{mn}^t is calculated by running the shortest cost algorithm in (N_b^t, A_b^t) and the cost to travel on each link $(i,j) \in A_b^t$ is $C_{ij}^{l_{ij}}$. Here, $C_{ij}^{l_{ij}}$ is the travel cost of each railcar on link $(i,j) \in A_b^t$ when l_{ij} trains have already traveled on that link. $(\rho_m^t \theta_m^t + r_b)C_m^C + \rho_n^t \theta_n^t C_n^C + \sum_{k \in P_{m\to n}-\{m\}-\{n\}} (1-\rho_k^t) r_b C_k^C$ is the additional classification cost along the train path, $P_{m\to n}$, for picking up block b at classification yard m and dropping the block at yard n.

If o_b(or d_b) is NOT in $N_b^t, d^t \neq o^t$, and $\sum_{i\in N^t}(1-\rho_i^t)+1 \leqslant R^{we}$, we extend the network of (N_b^t, A_b^t) to o_b (or d_b). We add o_b into N_b^t with a link from o_b to each node $n \in N_b^t$ into A_t^b if $1+\rho_n^t + X^t \leqslant R^{we}$, $\lambda_{o_b} + \rho_{o_b}^t \leqslant U_{o_b}^T, \lambda_n + \rho_n^t \leqslant U_n^T, \lambda_{o_b} + \rho_{o_b}^t \leqslant U_{o_b}^T$, $\gamma_{o_b} + \rho_{o_b}^t \theta_{o_b}^t \leqslant U_{o_b}^C$ and $\gamma_n + \rho_n^t \theta_n^t \leqslant U_n^C$ with the cost of $CN_{o_b n}^t$ as (31).

$$CN_{o_b n}^t = DC_{o^t n}^t r_b + TC_{o_b o^t}^t + \rho_n^t \theta_n^t C_n^C + C_{o^b}^C r_b + \sum_{k \in P_{o_b\to n}-\{o^b\}-\{n\}} (1-\rho_k^t) r_b C_k^C. \tag{31}$$

$DC_{o^t n}^t$ in (31) is the shortest path cost for train to taking the block from o^t to n and is similar to DC_{mn}^t in (30). The shortest path cost $TC_{o_b o^t}^t$ over links to extend train t's path so that it starts at o_b is calculated from o_b to o^t and picks block b from o_b by

- Updating the network of (N_b, A_b) after removing the links with U_{ij}^M trains and calculating the travel cost in each remaining link $(i,j) \in A_b$ as

$$C_{ij}^t = C_{ij}^{l_{ij}+1} r_b + \left(C_{ij}^{l_{ij}+1} - C_{ij}^{l_{ij}}\right)\varphi_{ij} + C_{ij}^T, \text{ and} \tag{32}$$

- Running the shortest cost algorithm from o_b to o^t in (N_b, A_b) to obtain $TC_{o_b o^t}^t$.

The remaining three terms are the additional classification costs for train t to take the block from o^b to n. By the above operations, o_b is added into N_b^t.

Similarly, the destination of the block d_b is added into N_b^t if the following conditions are met.

(1) $d_b \notin N_b^t$ and
(2) $\sum_{i\in N^t}(1-\rho_i^t)+1 \leqslant R^{we}$ if o_b was in N_b^t or $\sum_{i\in N^t}(1-\rho_i^t)+2 \leqslant R^{we}$ if o_b was not in N_b^t.

We add d_b into N_b^t with a link from each node $m \in N_b^t - \{o_b\}$ to d_b into A_b^t, if $1+\rho_m^t + X^t \leqslant R^{we}, \lambda_m + \rho_m^t \leqslant U_m^T, \lambda_{d_b} + 1 \leqslant U_{d_b}^T, \gamma_m + \rho_m^t \theta_m^t \leqslant U_m^C$ and $\gamma_{d_b} + \rho_{d_b}^t \theta_{d_b}^t \leqslant U_{d_b}^C$, with the cost of $CN_{md_b}^t$ as (33).

$$CN_{md_b}^t = DC_{md^t}^t r_b + TC_{d^t d_b}^t + (\rho_m^t \theta_m^t + r_b)C_m^C + \sum_{k \in P_{m\to d^t - \{m\}-\{d^b\}}} (1-\rho_k^t) C_k^C r_b. \tag{33}$$

$DC_{md^t}^t$ is calculated similar to DC_{mn}^t in (30) and also $TC_{d^t d_b}^t$ is calculated similar to $TC_{o_b o^t}^t$.

If both o_b and d_b are newly added nodes and $2+X^t \leqslant R^{we}, \lambda_{o_b}+1 \leqslant U_{o_b}^T, \lambda_{d_b}+1 \leqslant U_{d_b}^T \gamma_{o_b} + \rho_{o_b}^t \theta_{o_b}^t \leqslant U_{o_b}^C$ and $\gamma_{d_b} + \rho_{d_b}^t \theta_{d_b}^t \leqslant U_{d_b}^C$, we add a direct link from o_b to d_b into A_b^t with the cost of $CN_{o_b d_b}^t$, shown in (34).

$$CN_{o_b d_b}^t = \left(DC_{o_b o^t}^t + DC_{o^t d^t}^t + DC_{d^t d_b}^t + C_{o_b}^C\right) r_b + \sum_{k \in P_{o_b\to d_b}-\{o_b\}-\{d_b\}} (1-\rho_k^t) r_b C_k^C. \tag{34}$$

In Sep 5, a link from o_b to d_b is created as a new train with the cost of CN_{o_b,d_b}^{New} as (35) if $\lambda_{o_b} + \rho_{o_b}^t \leqslant U_{o_b}^T$ and $\lambda_{d_b} + \rho_{d_b}^t \leqslant U_{d_b}^T$ in (N_b^C, A_b^C). Please note that the $TC_{o_b d_b}^{New}$ is the minimum travel cost of routing a new train from o_b to d_b, which includes the possibly increased cost for the existing trains. The concept is the same as the way to calculate $TC_{o_b o^t}^t$ by using (32).

$$CN_{o_b d_b}^{New} = TC_{o_b d_b}^{New} + C^S + C_{o_b}^C r_b. \tag{35}$$

In Step 6 of BTA, all networks (N_b^t, A_b^t) for $t \in T^b$ are combined together into one network called $\left(N_b^C, A_b^C\right)$. For a link traveled by more than one train, the one with the least cost is selected. We create each link from o_b to $n \in N_b^C - \{d_b\}$ as a new train and add it into A_b^C with the cost of $CN_{o_b,n}^{New}$ based on (36), if $\lambda_{o_b} + \rho_{o_b}^t \leqslant U_{o_b}^T$ and $\lambda_n + \rho_n^t \leqslant U_n^T$ and $CN_{o_b,n}^{New}$ is less than the current cost of link (o_b, n) in $\left(N_b^C, A_b^C\right)$.

$$CN_{o_b n}^{New} = DC_{o_b n}^{New} r_b + C^S + C_{o_b}^C r_b. \tag{36}$$

Similarly, we create each link from $n \in N_b^C - \{o_b\}$ to d_b as a new train and add it into A_b^C with the cost of CN_{n,d_b}^{New} as (37), if $\lambda_n + \rho_n^t \leqslant U_n^T$ and $\lambda_{d_b} + \rho_{d_b}^t \leqslant U_{d_b}^T$ and CN_{n,d_b}^{New} is less than the current cost of link (n,d_b) in $\left(N_b^C, A_b^C\right)$.

$$CN_{n,d_b}^{New} = DC_{nd_b}^{New} r_b + C^S + C_n^C r_b. \tag{37}$$

The next step (Step 7) in the BTA algorithm is to find the shortest cost path from o_b to d_b in the network of (N_b^C, A_b^C). Each link in the shortest cost route represents a different train. Train information and network capacity information is updated after the assignment of a block. Please note that though the shortest-path algorithm is used frequently in BTA, the definition and the costs associated with those newly defined links in A_b^C are not the same as the physical links in the original network and the mileages of those physical links. Each newly-defined link (i,j) in $\left(N_b^C, A_b^C\right)$ indicates that block b is classified at yard j right after yard i. The congestion cost at links and classification cost at yards are considered in CN_{ij}^t if the block is carried by train t from yard i to yard j. These calculations are very different from the shortest-path algorithms used by MultiModal (Ireland et al., 2004) on the original physical network.

3.3. Train Routing (TR)

Each train made up in BTA with known start, termination, intermediate work event nodes is re-routed through an optimization model at the TR step. Similar to block sub-networks, we also create a train sub-network for each train. Before routing a train, *Pre-TR* steps are used to remove routing information of the train under consideration so that congestion effects are not double counted.

Pre-TR

Each train $t \in T$ carries a set of blocks $b \in B^t$ with known pickup and drop-off points (P_i^b, Q_i^b). The path of each train is known along with its X^t work event nodes and let N_p^t be the set of nodes in N^t with a working event, $\left(i.e. N_p^t = \bigcup_{b \in B^t} \left\{o_b^t, d_b^t\right\}\right)$. A train sub-network is created to reduce problem size by

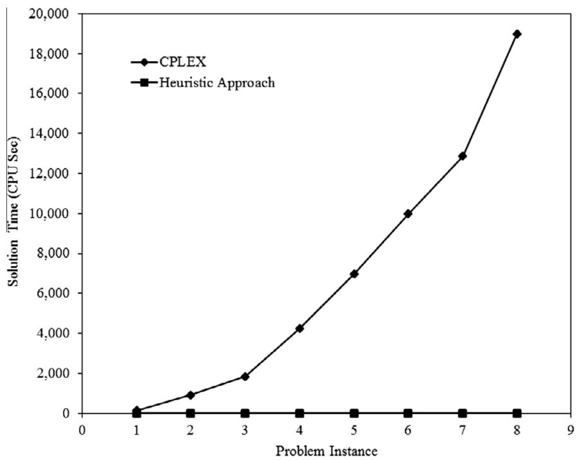

Fig. 5. Quality of heuristic solution approach based on solution time.

Table 2
Sample data of class-I railroad network: link attributes.

Origin ID	Destination ID	Distance (miles)	No. of Tracks	Type of Control	Capacity (feet)[a]	Capacity (tonnage)[a]	Capacity (trains)
0	1	40.1	1	CTC	4905	12870	30
0	2	62.4	1	TCS	5355	14220	30
1	2	58.7	2	CTC	4905	12870	75
1	3	67.4	1	ABS	3825	11880	18
2	3	69.8	1	TCS	4905	12870	30
2	4	31.0	1	ABS	4905	12870	18
2	5	25.9	1	CTC	4905	12870	30
3	5	48.0	1	TWC	4905	12870	16
3	6	21.0	1	TCS	5355	14220	30

[a] Capacity for each individual train traveling through the link.

Table 3
Sample data of class-I railroad network: node attributes.

Node ID	Capacity (cars)	Capacity (trains)
0	2000	48
1	1600	40
2	2200	55
3	1500	35
4	1200	30
5	2600	65
6	1500	42
7	1200	32
8	1800	42

- Adding node j into N^t if there is a pair of nodes (i,j) such that $i \in N^t, j \in N \backslash N^t$ and $D_{ij} \leqslant D$, where D is a predefined number (e.g. 100 miles);
- Adding any links that have both ends in N^t and are not currently in A^t into A^t;
- Deleting the links in A^t that have already reached the maximum train capacity restriction; and
- Letting l_{ij} and φ_{ij} be the numbers of trains and railcars, excluding train t, along link $(i,j) \in A^t$, where $l_{ij} = l_{ji}$ and $\varphi_{ij} = \varphi_{ji}$.

The TR optimization model is provided below with the following decision variable definition.

z_{ij}	whether the train travels along link (i,j), $(i,j) \in A^t$;
$u_i(v_i)$	whether the train starts (ends) at node $i, i \in N^t$;
a_i	whether the train starts and ends at node $i, i \in N^t$;
$g_i(h_i)$	length (tonnage) of the train after visiting node $i, i \in N^t$;
θ_i	number of railcars carried by the train after visiting node $i, i \in N^t$;
τ_{ij}	total additional car travel cost over link $(i,j) \in A^t$; and
k_i	an artificial variable to eliminate sub-tours, $i \in N^t$.

$$\text{Minimize} \sum_{(i,j)\in A_t} C_{ij}^{T} z_{ij} + \sum_{(i,j)\in A_t} \tau_{ij} \tag{38}$$

$$\text{S.T.}: \quad \sum_{\forall j:(i,j)\in A_t} z_{ij} - \sum_{\forall j:(j,i)\in A_t} z_{ji} = u_i - v_i \quad i \in N^t \tag{39}$$

$$\sum_{\forall j:(i,j)\in A_t} z_{ij} \leqslant 1 \quad i \in N^t \tag{40}$$

$$\sum_{\forall j:(i,j)\in A_t} z_{ji} \leqslant 1 \quad i \in N^t \tag{41}$$

$$\sum_{i\in N_t} u_i = 1 \tag{42}$$

$$2a_i \leqslant u_i + v_i \quad i \in N^t \tag{43}$$

$$k_j \geqslant k_i + 1 - Ma_j - (1 - z_{ij})M \quad (i,j) \in A^t \tag{44}$$

$$\sum_{\forall j:(i,j)\in A_t} z_{ij} + v_i \geqslant 1 \quad i \in N_p^t \tag{45}$$

$$k_j \geqslant k_i + 1 + M\left(P_i^b + Q_j^b - v_j - 2\right) \quad b \in B^t, i,j \in N^t \tag{46}$$

$$g_j \geqslant g_i + \sum_{b\in B_t} P_j^b L_b - (1 - z_{ij})M - \sum_{b\in B_t} Q_j^b L_b - M v_j \quad (i,j) \in A^t \tag{47}$$

$$(u_i - 1)M + \sum_{b\in B_t} P_i^b L_b \leqslant g_i \leqslant \sum_{\forall j:(i,j)\in A_t} U_{ij}^{L} z_{ij} \quad i \in N^t \tag{48}$$

$$h_j \geqslant h_i + \sum_{b\in B_t} P_j^b w_b - (1 - z_{ij})M - \sum_{b\in B_t} Q_j^b w_b - M v_j \quad (i,j) \in A^t \tag{49}$$

$$(u_i - 1)M + \sum_{b\in B_t} P_i^b w_b \leqslant h_i \leqslant \sum_{\forall j:(i,j)\in A_t} U_{ij}^{W} z_{ij} \quad i \in N^t \tag{50}$$

$$\theta_j \geqslant \theta_i + \sum_{b\in B_t} P_j^b r_b - (1 - z_{ij})M - \sum_{b\in B_t} Q_j^b r_b - M v_j \quad (i,j) \in A^t \tag{51}$$

$$\theta_i \geqslant (u_i - 1)M + \sum_{b\in B_t} P_i^b r_b \quad i \in N^t \tag{52}$$

$$\tau_{ij} \geqslant \left(C_{ij}^{l_{ij}+1} - C_{ij}^{l_{ij}}\right)\varphi_{ij} + C_{ij}^{l_{ij}+1}\theta_i - M(1 - z_{ij}) \quad (i,j) \in A^t \tag{53}$$

$$z_{ij}, u_i, a_i \in \{0,1\}; g_i, h_i, \theta_i, v_i, \varphi_{ij}, k_i \geqslant 0.$$

The objective function (38) minimizes the total cost, including train travel cost and railcar travel cost that depends on link volume, which is $C_{ij}^{l_{ij}+1}$. Constraint set (39) keeps the flow conversation. Constraint sets (40) and (41) make sure that train t can visit each node (link) at most once. Constraint set (42) ensures that the train can start only once while constraint set (43) permits the train to start and end at the same node. Sub-tours are eliminated by constraint set (44). Constraint set (45) guarantees that the train definitely leaves the node where the train has a work event, except the termination node. The train must pick up a block before dropping the block, guaranteed by constraint set (46) by using the variables ordering nodes visited by the train from (44). Please note that constraint set (46) is always loose when the dropping point is the termination node. Length restriction of the train over each link is ensured by constraint sets (47) and (48). Similarly, the weight restriction over each link is met by constraint sets (49) and (50). The number of railcars the train has after visiting a node is obtained by constraint sets (51) and (52). Finally, the additional railcar traveling cost caused by routing this train over a link is obtained by constraint set (53). Though the model (38)–(53)seems complicated, it is much smaller than the original model (1)–(29) because it has a much smaller network and is only for one train. Most of variables in the original model (1)–(29)become known (parameters) with values from BTA. The model for each train can be solved to its optimum with commercial solvers, such as IBM CPLEX. However, we note that the resulting solution may be suboptimal to the overall train routing problem.

3.4. Numerical experiments to show the computational effectiveness of the heuristic method

The iterative heuristic algorithm is implemented using C++ and calling IBM CPLEX 12.3 for the TR sub-problem on a PC with 3.30 GHz CPU and 16 GB RAM. The MIP model (1)–(29) of the overall problem is also solved to optimality for small instances by CPLEX 12.3 as a benchmark. Please note the optimization model (1)–(29) cannot be solved for large-scale instances. Table 1 summarizes the results for ten randomly generated small-sized instances. The second column reports the problem size in terms of the numbers of nodes, links, and blocks. The optimization model (1)–(29) could not be solved to its optimality by CPLEX within the 20,000-s time limit from Instance 9. The heuristic algorithm provides quality solutions with much higher speed (less than 1 s). For the first small instance, the heuristic algorithm reaches the same optimal solution from CPLEX. Table 1 justifies the superiority of the proposed heuristic algorithm regarding both solution quality and computational time. The computational time of the proposed heuristic algorithm is also shown in Fig. 5.

4. Case study

The proposed solution approach is applied to a Class-I railroad network to evaluate the criticality of its network infrastructures. The network consists of 200 nodes and 478 links. The case study evaluate the criticality of each element, a node or a link, by comparing the total travel cost (time) before and after the disruption to (removal of) this element. The study uses two types of data: network data and freight flow data. The flow data contain the 552 blocks' ODs and their attributes (e.g., number of cars, lengths and tonnages). The network attributes include distance between stations/yards, number of tracks, and track operating characteristic (e.g., signal types). The base reference is the software North American Railroad Map, Version 3.11, which reflects the North American railroad network in 2010. Using the track attributes, we estimated the capacity in the number of trains per day based on Cambridge Systematics (2007) and other parameters (e.g., length and tonnage restrictions allowed for each traveled train) of each link. Information available at the railroad's website aided this process.

Table 4
Sample freight flow data.

Block ID	Origin ID	Destination ID	No. of Cars	Length (feet)	Weight (tonnage)
0	0	1	25	1362	3575
1	1	8	10	545	1430
2	0	7	16	873	2288
3	2	6	25	1362	3575
4	3	5	12	655	1716
5	4	6	19	1035	2717

Table 5
Other input parameters for the case.

Parameter	Value
Train starting cost ($)	400
Train travel cost ($/mile)	10
Car classification cost ($)	40
Car travel cost without congestion ($/mile)	0.5
Car travel cost at moderate congestion ($/mile)	0.75
Car travel cost at high congestion ($/mile)	1.5
Maximum blocks per train	8
Maximum number of work events per train	6

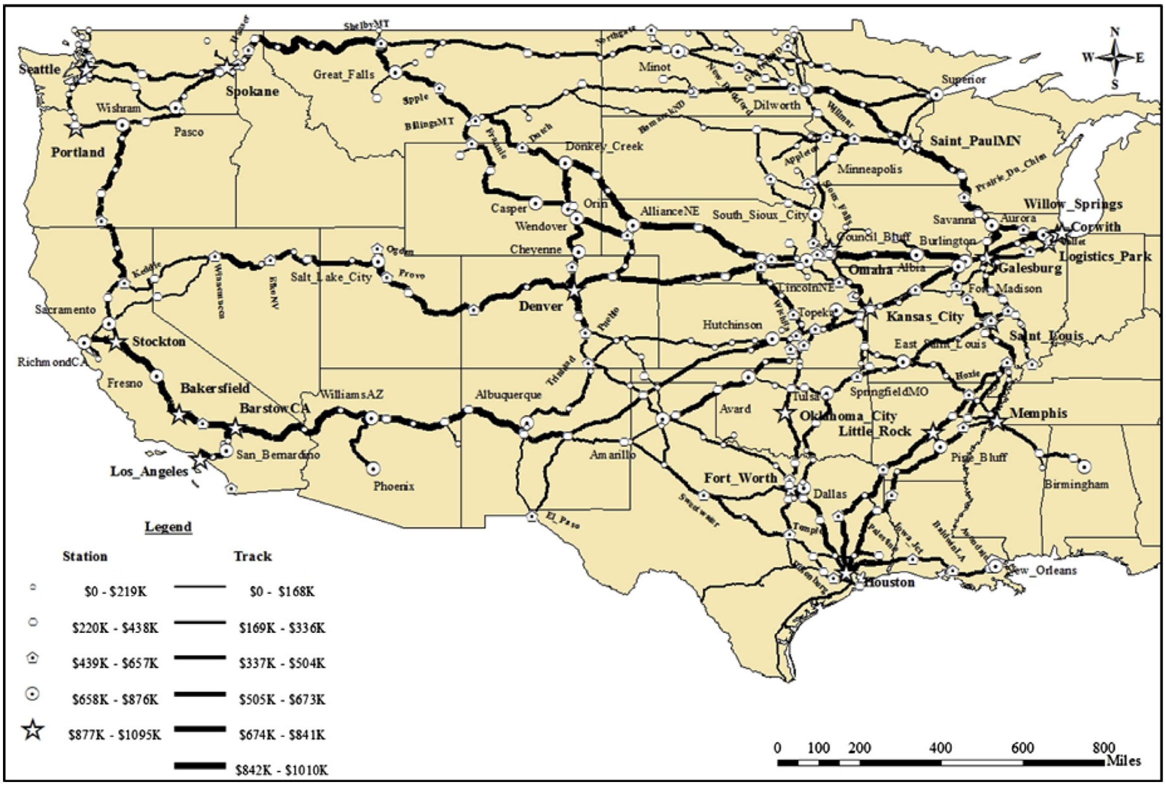

Fig. 6. Criticality map of one Class-I railroad network.

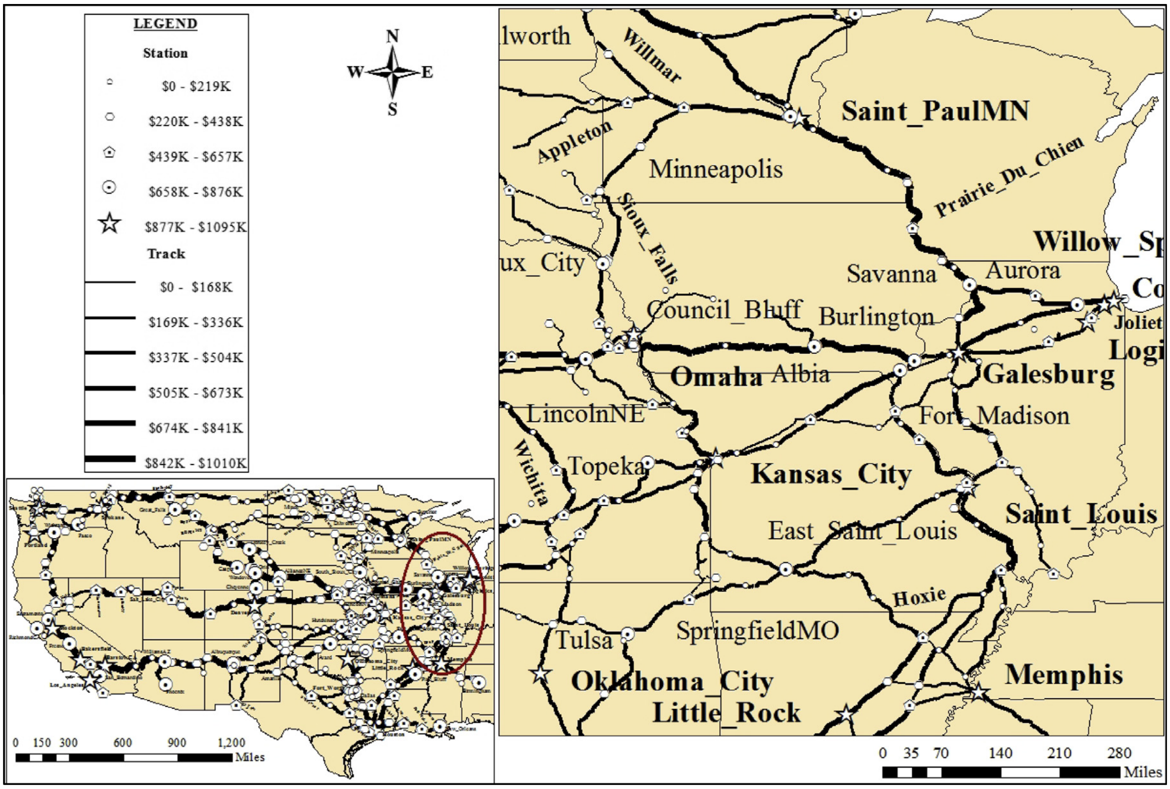

Fig. 7. Criticality map in more details.

Station/Yard capacities (Marinov and Viegas, 2009) were derived from the railroad's websites and correspondence with railroad personnel. Tables 2 and 3 provide sample network data for links and nodes, respectively.

Freight flow data were collected from the Freight Analysis Framework (FAF) database. It combines data from several sources to create a comprehensive image of freight movement among states and major zones by all transportation modes. FAF version 3 provides the most recent freight flow data of tonnage and value, by commodity type, mode and OD. This study uses 2009 rail freight data for all OD pairs by tonnage from (FHWA, 2009). Blocks per day for each OD are calculated based on (Cambridge Systematics, 2007). The sample freight flow data used in the case study is given in Table 4. Table 5 shows various cost parameters and train attributes collected from (Cambridge Systematics, 2007; Ahuja et al., 2005). Three levels of congestion: normal (no congestion), moderate, and high are considered at links.

The proposed heuristic solution approach, outlined in Fig. 1, is applied to the whole network to obtain the total cost, which is considered the 'standard' transportation cost. Subsequently, each network element, link or node, is considered to be disrupted *individually*(removed from the network) and all traffic is re-routed in the residual network with the heuristic approach. The criticality of each element is estimated by the increased transportation cost caused by the disruption to that element compared to the standard transportation cost. Fig. 6 demonstrates the criticality levels of all elements for this case, which is obtained by running the heuristic approach for about 900 times, one for the disruption of each element (link or node). Each run takes about 15 min. Fig. 7 shows the enlarged views of the criticality levels of the part of network infrastructures for better visibility and legibility. The relative criticality level of a link or yard is represented by its width or size. Wider links or bigger yards are more critical. A disruption to those elements will result in more transportation cost (delay) for the whole rail freight flows. The stakeholders of rail network should pay more attention to protect or add redundancy to those elements in order to enhance the whole network's resilience.

5. Conclusion and future work

This paper evaluates the criticality of railroad network components by optimally creating and routing of trains through a disrupted and possibly congested network. Because of the expected congestion and capacity constraints, existing train forming and routing models and algorithms in the literature, which were developed for normal operations, could not be used under an event of disruption. This paper proposes an optimization model considering the capacity at both links and yards and the speed–volume relationship at links. An iterative heuristic solution method is proposed to solve the large-scale instances. The numerical study based on small instances shows that the heuristic solution approach can yield high quality solutions with high speed. The acceptance of computational speed is further verified by the case. Currently, we were told by several Class-I railroads that the freight rail network in US are in general running under its capacity. Both the model and the solution method proposed by this paper may not be appropriate for normal operational planning because the capacity and congestion concerns at links and nodes do significantly increase the complexity of the model. In fact, the normal train forming and routing may have to consider other factors, such as crew scheduling and routing, track rights, locomotive availability, timetables, etc. However, the methodology described in this paper may help them develop long-term strategic plans and develop preparedness and response plans for disruptions based on criticality level of infrastructure. Criticality of each individual link or yard is measured by the increased transportation cost (delay) after re-routing the trains through the residual network in the absence of the element. Another limit of this research is that the application of the proposed methodology in practice depends on capacity models on links and yards. Several on-going efforts to study railroad capacity supported by Transportation Research Board and the NuRail Center[2] are expected to improve the applicability of this proposed method.

The method could also be used to estimate the effect of a certain disruption scenario that may involve multiple rail infrastructure elements. However, the paper does not demonstrate this application because of the large number of possible scenarios and therefore large computational burden. Our methodology for evaluating the criticality might be improved by developing a decision support system (DSS) based on this heuristic algorithm. It should have the feature to arbitrarily disrupt multiple network components and run the model to evaluate their criticality. The DSS may have more practical value if interfaced with the commercial geographic information system (GIS) software.

Acknowledgment

This research was partially supported by the US Department of Transportation, Office of the Secretary, Grant No. DTOS59-09-G-00058.

References

American Association of State Highway and Transportation Officials (AASHTO), 2007. Transportation – Invest in our future: America's freight challenge.
Ahuja, R.K., Cunha, C.B., Sahin, G., 2005. Network models in railroad planning and scheduling. INFORMS 2005 Tutorials in Operations Research, 54–101.
Ahuja, R.K., Jha, K.C., Liu, J., 2007. Solving real-life railroad blocking problems. Interface 37 (5), 404–419.
Assad, A.A., 1980a. Modeling of rail networks: toward a routing/makeup model. Transportation Research Part B 14 (1-2), 101–114.
Assad, A.A., 1980b. Models for rail transportation. Transportation Research Part A 14 (3), 205–220.
Assad, A.A., 1981. Analytical models in rail transportation: an annotated bibliography. INFOR 19 (1), 59–80.

[2] http://www.nurailcenter.org/

Barnhart, C., Jin, H., Vance, P.H., 2000. Railroad blocking: a network design application. Operations Research 48 (4), 603–614.
Brannlund, U., Lindberg, P.O., Nou, A., Nilsson, J.E., 1998. Railway timetabling using Lagrangian relaxation. Transportation Science 32 (4), 358–369.
Cacchiani, V., Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L., Wagenaar, J., 2014. An overview of recovery models and algorithms for real-time railway rescheduling. Transportation Research Part B 63 (1), 15–37.
Cambridge Systematics, Inc., 2007. National Rail Freight Infrastructure Capacity and Investment Study. Association of American Railroads.
Caprara, A., Kroon, L., Toth, P., 2011. Optimization problems in passenger railway systems. Wiley Encyclopedia of Operations Research and Management Science.
Carey, M., 1994. A model and strategy for train pathing with choice of lines, platforms, and routes. Transportation Research Part B 28 (5), 333–353.
Colombo, F., Cordone, R., Trubian, M., 2011. A column-row generation heuristic for the train design optimization problem. Presented at INFORMS 2011 Annual Meeting – RAS 2011 Problem Solving Competition, Charlotte, NC.
Cordeau, J.F., Toth, P., Vigo, D., 1998. A survey of optimization models for train routing and scheduling. Transportation Science 32 (4), 988–1005.
Davis, S.C., Diegel, S.W., Boundy, R.G., 2012. Transportation Energy Data Book: Edition 32.
Eakin, K., Meitzen, M., Bozzo, T., Schoech, P., 2009. A Study of Competition in the U.S. Freight Railroad Industry and Analysis of Proposals that Might Enhance Competition. Surface Transportation Board.
Farvolden, J.M., Powell, W.B., 1994. Subgradient methods for the service network design problem. Transportation Science 28 (3), 256–272.
Federal Highway Administration (FHWA), 2009. Analysis Framework Version 3. <http://faf.ornl.gov/fafweb/Extraction1.aspx>. Accessed on January 9, 2014.
Gorman, M.F., 1998. An application of genetic and Tabu searches to the freight railroad operating plan problem. Annals of Operations Research 78 (1), 51–69.
Haghani, A.E., 1987. Rail freight transportation: a review of recent optimization models for train routing and empty car distribution. Journal of Advanced Transportation 21 (2), 147–172.
Haghani, A.E., 1989. Formulation and solution of a combined train routing and makeup, and empty car distribution model. Transportation Research Part B 23 (6), 433–452.
Ireland, P., Case, R., Fallis, J., Van Dyke, C., Kuehn, J., Meketon, M., 2004. The Canadian Pacific Railway transforms operations research by using models to develop its operating plans. Interfaces 34 (1), 5–14.
Jha, K.C., Ahuja, R.K., Sahin, G., 2008. New approaches for solving the block-to-train assignment problem. Networks 51 (1), 48–62.
Kaspi, M., Raviv, R., 2013. Service-oriented line planning and timetabling for passenger trains. Transportation Science 47 (3), 295–311.
Jin, J.G., Zhao, J., Lee, D.H., 2013. A column generation based approach for the train network design optimization problem. Transportation Research Part E 50 (1), 1–17.
Kwon, O.K., Martland, C.D., Sussman, J.M., 1998. Routing and scheduling temporal and heterogeneous freight car traffic on rail networks. Transportation Research Part E 34 (2), 101–115.
Keaton, M.H., 1989. Designing optimal railroad operating plans: Lagrangian relaxation and heuristic approaches. Transportation Research Part B 23 (6), 415–431.
Keaton, M.H., 1992. Designing optimal railroad operating plans: a dual adjustment method for implementing Lagrangian relaxation. Transportation Science 26 (4), 263–279.
Lai, Y.-C.(Rex), Barkan, C.P.L., 2009. Enhanced parametric railway capacity evaluation tool. Transportation Research Record: Journal of the Transportation Research Board 2117, 33–40.
Lozano, L., González, J.E., Medaglia, A.L., 2011. A network-oriented formulation and decomposition approach to solve the 2011 RAS problem. Presented at INFORMS 2011 Annual Meeting – RAS 2011 Problem Solving Competition, Charlotte, NC.
Marín, A., Salmerón, J., 1996. Tactical design of rail freight networks. Part I: Exact and heuristic methods. European Journal of Operational Research 90 (1), 26–44.
Marinov, M., Viegas, J., 2009. A simulation modeling methodology for evaluating flat-shunted yard operations. Simulation Modeling Practice and Theory 17 (6), 1106–1129.
Newman, M., Nozick, L., Yano, C.A., 2002. Optimization in the rail industry. In: Pardalos, P.P., Resende, M.G.C. (Eds.), Handbook of Applied Optimization. Oxford University Press, New York.
Newman, A.M., Yano, C.A., 2001. Scheduling trains and containers with due dates and dynamic arrivals. Transportation Science 35 (2), 181–191.
Newton, H.N., Barnhart, C., Vance, P.M., 1998. Constructing railroad blocking plans to minimize handling costs. Transportation Science 32 (4), 330–345.
Petersen, E.R., 1977. Rail yard modeling: Part II. The effect of yard facilities on congestion. Transportation Science 11 (1), 50–59.
Shih, M-C., Dick, C.T., Sogin, S.L., Barkan, C.P.L., 2014. Comparison of capacity expansion strategies for single-track railway lines with sparse sidings. Transportation Research Record: Journal of the Transportation Research Board, to Appear.
Stodolsky, F., 2011. Railroad and Locomotive Technology Roadmap. U.S. Department of Energy.
Transportation Security Administration (TSA), 2012. Freight Rail. <http://www.tsa.gov/what_we_do/tsnm/freight_rail/index.shtm>, Accessed June 30, 2012.
Turnquist, M/A., Daskin, M.S., 1982. Queuing models of classification and connection delay in railyards. Transportation Science 16 (2), 207–230.
Wang, I.-L., Lee, H.-Y., Liang, Y.-T., 2011. Train design optimization. Presented at INFORMS 2011 Annual Meeting – RAS 2011 Problem Solving Competition, Charlotte, NC.
Yoon, H., Cheng, C., Dimri, A., 2011. Meet-pass planning model for rail capacity analyzer. Presented at INFORMS 2013 Annual Meeting, Minneapolis, MN.
Yue, Y., Zhou, L., Yue, Q., Fan, Z., 2011. Multi-Route railroad blocking problem by improved model and ant colony algorithm in real world. Computers & Industrial Engineering 60 (1), 34–42.