Multimodal Dynamic Freight Load Balancing

Afshin Abadi, Petros Ioannou, Fellow IEEE, and Maged M. Dessouky

Abstract—The urban traffic network has temporal and spatial characteristics whose changing conditions have often unpredictable effects on the flow of loads that include passengers and freight. As a result, the current traffic network is unbalanced leading to high and low peaks of traffic in both time and space. The freight transportation chain can utilize these high and low peaks in the road and rail network in order to utilize more effectively available capacity.

The purpose of this paper is to develop a coordinated Multimodal Dynamic Freight Load Balancing (MDFLB) system to balance freight loads across the road and rail network. The MDFLB system collects and updates information from all the shipping companies and assigns freight loads to the available carriers using an optimization model, while taking into account current and predicted dynamical changes in the associated networks. Since the freight loads can change the assumed states of the network, namely the link travel times, which could render the solution of the optimization problem no longer optimal, an iterative approach is considered involving online network simulation models. The simulation models are used to test and modify the optimization-based load balancing solution and estimate the new states of the network used by the optimizer. This feedback iterative approach guarantees that the overall cost function is non-increasing and it stops when it converges to a minimum or when a stopping criterion is satisfied depending on the time horizon of interest. A simulation case study that simulates on distribution of freight in an area that includes the two major sea ports in Southern California is used to demonstrate the effectiveness of the proposed coordinated MDFLB.

Index Terms—Simulator, traffic congestion, optimization, freight.

I. INTRODUCTION

The routing of goods involves complex operations in a dynamical network with uncertainties and limited resources. This complex system leads to congestion and delays that have significant impact on the quality of life and the environment as well as the economy. Some of the usual causes of congestion are disruptions and rapid changes in demand. Disruptions such as road accidents, emergency road closure etc. are unforeseen. In the case of pre-planned events such as cultural events, security checkpoints, routine maintenance, or road closure, disruptions can be predicted ahead of time. Sea-ports often located adjacent to an urban area with time dependent traffic generate demand for freight transport that has a major impact on roadway traffic and congestion. These disruptions often generate a lot of uncertainties that lead to inefficient balancing of loads across the network. Furthermore, the lack of cooperation between the different users of the system adds to the complexity of the problem.

Distribution systems offer the potential of aggregation and sharing of multiple resources. These resources could be owned and operated by different organizations. Each individual organization may have different objectives, goals, and strategies from other organizations. Load balancing in a distribution system is a process of balancing workloads by taking into account the organization objectives. The load balancing process can be either static or dynamic. In the static load balancing process, statistical behavior of the network is considered and transfer decisions are made without taking into account the current and/or predicted system state. Transfer decisions involve volumes and the best available routes to transfer loads from origins to destinations. Such decisions are not based on the dynamic nature of the network and thus the solution may be far from optimal if the system states deviate from the assumed static conditions. On the other hand, a dynamic load balancing system takes into account historical, current, and predicted states. The main advantage of the static load balancing is its simplicity as it can be performed offline; whereas, in the dynamic load balancing case the current and predicted system states must be fed to the solution procedure in real time. However, dynamic load balancing has the potential of providing better solutions for the distribution system. The dynamic load balancing process can be performed coordinated or uncoordinated. In the uncoordinated case each user chooses how to balance its load across the networks based on the information available by minimizing its own cost objective. It has no information about the load distribution of other users and they will affect its objective. In the coordinated case each user sends its load demand to a central coordinator who then provides optimum routes to each user. Therefore given the additional information available the coordinated load balancing approach is expected to provide lower overall cost. An association which receives information from shipping companies, freight forwarders, transport operators, and the port authority and provides the best feasible routes to transport around the port areas in Southern California is an example of a coordinator.

The solution obtained from the coordinated dynamic load balancing works for the overall system but it may not satisfy an individual distributor system’s goals. However, uncoordinated dynamic load balancing may satisfy the individual distributor system’s objective at possibly the expense of the overall system’s performance. Another issue with coordinated versus uncoordinated is access to network information and overall demand. A coordinated system may have more accurate information than an uncoordinated one and therefore it could lead to better solutions for the overall system and individual distributor too. The question whether optimality for the overall system based on full network information is also a better solution for the individual distributors with less network information is an important research topic. If it is a worse solution for some distributors and better for others then the issue of fairness and providing incentives for participation for the individual distributors who may be worse off an important question. In this paper we focus only on the development and evaluation of a coordinated MDFLB system and leave the other topics for future research.

The main contribution of this paper is the development and evaluation of a coordinated MDFLB using a co-simulation optimization approach. The approach employs a macroscopic traffic flow simulator and a railway simulation system to find the...
The approach is based on the fact that an initial load balancing solution may change the link states of the network, which may destroy the optimality of the initial solution. The simulation models are used to update the link states and calculate the new link costs. The updated costs are used to generate an updated optimum solution and the process is repeated in an iterative feedback manner till convergence is achieved or a stopping criterion is satisfied.

This paper is organized as follows: A review of relevant literature on the subject is presented in Section II. The problem description and the simulation model framework is described in Section III. The results of the computational experiments based on the proposed methodology are presented in Section IV. The Port of Los Angeles (LA) and Long Beach (LB) together with the associated rail and road network are considered as the region of study. The conclusions are included in Section V.

II. LITERATURE REVIEW

Numerous papers have been published to address multimodal freight transportation. Formulation and solution of a multimodal network flow model were presented by Haghani et al. [1]. The multimodal network was based on a time-space network, and two heuristics were proposed to solve the problem. Jourquin et al. [5] analyzed a multimodal freight transportation policy in Europe. They introduced a model for freight moving, loading, and unloading. The proposed optimization algorithm aimed to minimize the total transportation cost including route choices, modes, and means. Modesti et al. [3] provided a utility measure for finding multi-objective shortest paths in urban multimodal transportation networks. They presented an approach based on the classical shortest path problem for a multimodal transportation network. The experiments were performed on the urban transportation network of an Italian city. Lozano et al. [4] extended their previous work by using label correcting techniques to find the shortest viable path in a multimodal transportation network.

Intermodal and international freight network modeling was studied by Southworth et al. [5]. They constructed a network that covered five million origin-to-destination freight shipments reported as part of a 1997 United States commodity flow survey. Multimodal transportation, logistics, and the environment were studied by Rondinelli et al. [6]. They examined the interaction among the transportation activities and the types of environmental impact emanating from multimodal transportation operations. Supply-demand equilibrium in a multimodal transportation network was proposed by Fernandez et al. [7]. They developed a new approach for an intercity freight transportation system that takes into account the supply-demand equilibrium. Arnold et al. [8] presented an approach for modeling a rail/road intermodal transportation system. The experiments were focused on the rail/road transportation system in the Iberian Peninsula. Macharis et al. [9] provided a comprehensive review of an intermodal freight transportation system. They argued that intermodal freight transportation research is emerging as a new research application field which needs different types of models than those applied to uni-modal transport system. Also, Nijkamp et al. [10] presented comparative modeling of interregional transport flows which was applied to multimodal European freight transport. They compared the descriptive and predictive power of two classes of statistical estimation models for multimodal network flows.

Scheduling multimodal freight transportation systems is an important issue which has been addressed by many researchers. Castelli et al. [11] used a Lagrangian based heuristic procedure for scheduling multimodal transportation networks. At each step, their proposed algorithm schedules a single line by correcting the previous decisions. Ham et al. [12] provided implantation and estimation of a combined model of interregional, multimodal commodity shipment and transportation network flows. They described the formulations and solutions of the model using U.S. interregional commodity shipment data, and evaluated the model with the observed data. Optimizing the design of a multimodal freight transport network in Indonesia was proposed by Russ et al. [13]. A mathematical model is developed within the framework of a bi-level programming problem, where a multimodal multi-user assignment forms the lower level problem and the upper level includes the combination of actions for capacity expansion.

Environmental issues are taken into account in multimodal freight transportation system modeling. Janic [14] developed an approach for modeling the full costs of a multimodal and road freight transport network. A heuristic approach was presented by Yamada et al. [15] for designing a multimodal freight transportation network. The proposed model determines a suitable set of actions for improving the existing infrastructure or establishing new railways, roads, and freight terminals. A mathematical model of selecting transport facilities for multimodal freight transportation is provided by Lingaitiene [16]. The cost function takes into account the overall technological costs of transportation using road, rail, and sea links. Caris et al. [17] studied planning problems in multimodal freight transportation systems. They extended their previous work by proposing decision support models for network, drayage, and terminal operators as well as for public actors such as policy makers and port authorities. Many other publications that addressed relevant multimodal freight transportation systems can be found in references [18]–[25].

However, none of the mentioned prior work addresses the coordinated Multimodal Dynamic Freight Load Balancing (MDFLB) problem where the routing of freight is done in a coordinated manner by taking into account the overall benefit. Moreover, the mentioned works do not take into account the dynamic nature of traffic conditions which is non-linear with respect to the volume and is time dependent. As mentioned earlier, for a certain range of flow, the cost of using a link and consumption of capacity may change linearly, but at some break point the congestion increases as a nonlinear function of the volume. To overcome these limitations, we formulate the optimization problem using the framework of a service network design problem [26] with an appropriate cost function to find the optimal solution for the coordinated MDFLB. We use a real time simulation model of the traffic on the road and rail networks in order to capture dynamic behaviors and predict the cost terms used in the optimization model.

III. PROBLEM DESCRIPTION

The transportation network can be represented as a graph consisting of a set of nodes with arcs connecting the nodes. Nodes ($N$) represent origins ($n_o$), intermediary ($n_i$), and destinations ($n_d$). A set of arcs in the network is characterized by
the set of available transportation links $L$ (roads/railways) between 62
origins and destinations.

Let $X(k) = [x_1(k) \ldots x_L(k)]'$ be a vector with the volumes of the 63
links 1 to $|L|$ at time $k \in K$ where $K = \{0, 1, \ldots, T\}$ and $T$ 64
represents the time horizon of interest. Let $d_i(k)$ denote the 65
demand or supply of node $i$ at time $k$. In this formulation, we do 66
not assume a specific origin point satisfies a destination point. 67
That is, demand is not specified as an origin-destination pair but 68
that any one of the several supply nodes $i$ can satisfy any demand 69
node. The historical link volumes $\{X(0), \ldots, X(k)\}$ and future 70
demands or supplies of node $i$ $\{d_i(k+1), \ldots, d_i(k+m)\}$ are 71
assumed to be available. The amount of supply or demand of 72
node $i$ at time $k$ can be defined as follows:

\[
\begin{align*}
\text{supply} &= d_i(k) & \text{if} \ i \in n_0 \\
\text{demand} &= -d_i(k) & \text{if} \ i \in n_D \\
d_i(k) &= 0 & \text{if} \ i \in n_T
\end{align*}
\]

The relationship of the link volumes with $d_i(k)$ and other 77
parameters in the network can be expressed by the nonlinear 78
dynamical equation.

\[
X(k + 1) = X(k) + f(X(k), k, p(k), A_i(k)d_i(k))
\]

where $A_i(k) = [a_{i1}(k) \ldots a_{iL}(k)]'$ is an array which determines 89
the proportions of $d_i(k)$ which are assigned to the links 1 to $|L|$ 90
from node $i$. $f(X(k), k, p(k), A_i(k)d_i(k))$ is a non-linear function 91
of the link volumes, time, the impact of the other vehicles from 92
adjacent links denoted by $p(k)$, and $A_i(k)d_i(k)$ which is the 93
assigned number of units to the links from node $i$. The impact of 94
the other vehicles on the link $(p(k))$ is the number of vehicles at 95
time $k$ that is estimated to be on the link at time $k + 1$ from the 96
adjacent links. The link volumes in the network are time-dependent 97
and change due to various reasons such as added load and 98
incidents. The analytical form of model (2) in terms of equations 99
is complicated due to the non-linearities and complex interactions 100
between the variables. These non-linearities and interactions 101
however can be easily programmed using commercially available 102
traffic simulators. Therefore instead of using an overly simplified 103
version of (2) to generate the link volumes a simulator can be used 104
to generate far more accurate link volumes. Our approach is based 105
on incorporating an online simulator in generating the link 106
volumes which are then used to generate or predict the costs along 107
the links to be used in an optimization procedure.

Let $C(k + 1) = [c_1(k + 1) \ldots c_L(k + 1)]$ be an array which 118
specifies the total costs per unit for the links 1 to $|L|$ at time $k+1$. 119
The cost $C(k + 1)$ is calculated using the volumes of links $X(k + 1)$ 120
generated by the computer simulation model and other 121
parameters depending on the objective of the individual distributor 122
system. The uncoordinated dynamic load balancing model aims 123
find the best solution based on the information of an individual 124
distributor system. Hence, the uncoordinated dynamic load 125
balancing problem for origin $i$ can be formulated as follows:

\[
\begin{align*}
\text{minimize}_{A_i} & \sum_{k=0}^{T-1} C(k + 1)A_i(k)d_i(k) \\
\text{s.t.} & \quad X(k + 1) = X(k) + f(X(k), k, \check{p}(k), A_i(k)d_i(k))
\end{align*}
\]

where $u_i$ denotes the maximum capacity of link $l \in L$; $\check{p}(k)$ is 140
the estimated impact of other vehicles assumed by the individual 141
distributor. Equation (8) indicates available information in msteps 142
in the future at time step $k$ and $k + m \leq T$.

The objective function (3) aims to minimize the total cost of the 153
network of the individual distributor. As mentioned earlier, the 154
cost $C(k + 1)$ is calculated using the volumes of links $X(k + 1)$ 155
generated by the computer simulation model based on the assigned 156
number of units from node $i$ at time $k$ $(A_i(k)d_i(k))$ and the other 157
parameters. We use time $k$ for the quantity $(A_i(k)d_i(k))$ and $k+1$ 158
for $C(k+1)$ since we assume a time lag of one unit between making 159
the assignment to the link and the actual travel on the link. 160
Constraint (4) generates the link volumes, constraint (5) 161
corresponds to the maximum volume of the links, and constraint 162
(6) represents the flow conservation at node $i$ where $Q_{1i}$ is a set 163
of links originating from node $i$ and $Q_{2i}$ is a set of links ending 164
at node $i$.

In the coordinated dynamic load balancing, a coordinator is 175
responsible to collect and update the information from all 176
distribution systems in order to make a final transfer decision. 177
Therefore, the coordinated dynamic load balancing can be 178
presented as follows:

\[
\begin{align*}
\text{minimize}_{A_1 \ldots A_N} & \sum_{k=1}^{T} \left[ C(k + 1) \sum_{i=1}^{N} A_i(k)d_i(k) \right] \\
\text{s.t.} & \quad X(k + 1) = X(k) + f\left(X(k), k, \check{p}(k), \sum_{i=1}^{N} A_i(k)d_i(k)\right) \\
& \quad 0 \leq x_i(k + 1) \leq u_i & \forall i \in L \\
& \quad \sum_{l \in Q_{1i}, j \in N} a_{lj}(k)d_j(k) - \sum_{l \in Q_{2i}, j \in N} a_{lj}(k)d_j(k) = d_i(k) & \forall i \in N \\
given & \quad \{X(0), ..., X(k)\} \\
and & \quad \{d_i(k+1), ..., d_i(k+m)\}
\end{align*}
\]

Equation (10) represents the dynamic nonlinear characteristics 188
of the link volumes in response to new freight loads and passenger 189
traffic. The specific form and parameters of the non-linear function 190
$f$ are unknown. One approach is to identify the parameters and 191
nonlinear terms of the function $f$ in a closed form and then use it 192
to estimate the link volumes. Such a task is very complicated if at
The proposed methodology starts with initial cost estimates of the link routes based on historical data and simulation models, which are used by the optimization to find the minimum cost route. The load balancing block adds loads to the minimum cost route until its costs are the same as that of the other routes with freight loads. Since the loading will change the states of the network it will lead to possibly new cost estimates in which case the optimization algorithm is exercised again to lead to a new minimum cost route which then the load balancing block has to balance by shifting loads from higher cost routes to the minimum cost one. This iterative process continues until one of the following stopping criteria is satisfied:

1. The changes in the total cost is less than a predefined value in two consecutive iterations.
2. The maximum number of iteration is reached.

In each step of the procedure the cost function decreases as the selected lower cost route is reduced in cost by shifting loads to unselected higher cost routes. Therefore, the cost is a nonlinear increasing function with respect to iterations (see also example below). Since the cost function is also bounded from below by zero, its convergence is guaranteed by using basic results from real analysis. The new states of the network are fed back to the optimizer which may generate new minimum cost routes which indicates repeating the procedure of load balancing for the new routes.

Once the stopping criterion is satisfied the final load balancing solution is applied to the actual system. Fig. 2 represents the algorithm procedure. In the following section we discuss in detail the methodology of Fig. 1.

**Fig. 1. Framework of the proposed methodology**

**Fig. 2. The algorithm procedure**

**Optimization**

The objective of the optimization formulation is to select the input volume sequence and distribution of freight on the links in the network in order to satisfy customer demands and constraints at minimum system cost. Let $D_i$ represent the total amount of supply or demand of node $i$ which can be defined as follows:

$$\begin{align*}
\text{supply} = D_i & \quad \text{if } i \in n_0 \\
\text{demand} = -D_i & \quad \text{if } i \in n_D \\
D_i = 0 & \quad \text{if } i \in n_I
\end{align*}$$

(15)

Note that $\sum_{i \in N} D_i = 0$ and nodes consist of origins ($n_0$), intermediary ($n_I$), and destinations ($n_D$). In the coordinated MDFLB, the coordinator collects and updates information from all the origins in the network. Therefore, the optimization formulation for the coordinated MDFLB can be expressed as (9)-(14) with the additional constraint.

$$\sum_{k \in K} d_i(k) = D_i \quad \forall i \in n_0$$

(16)

Note that in the optimization formulation, the cost vector $C(k + 1)$ associated with the various links is evaluated using a nonlinear function of the link volumes $X(k + 1)$. In practice the congestion increases as a nonlinear function of the volume, and we capture these nonlinearities in the objective function using a simulation model and cost evaluators that will be explained in the next section. The iterative procedure of Fig. 1 aims to find the final solution from a finite set of possible solutions because there exist a finite set of routes from the origins to destinations. We solve the optimization model using MATLAB.

We present two simple examples to illustrate the iterative procedure of Fig. 1. Let us assume a network with 5 nodes including origin ($n_0$), destination ($n_D$), and three intermediary nodes ($n_I$). There exist three possible routes from $n_0$ to $n_D$ as shown in Fig. 3.

**Fig. 3. Simple network**
For the first example, it is assumed that all the links have identical characteristics. The current traffic conditions (current demands) on the links are presented in Table I.

<table>
<thead>
<tr>
<th>Link</th>
<th>((n_0, n_1))</th>
<th>((n_2, n_3))</th>
<th>((n_4, n_5))</th>
<th>((n_6, n_7))</th>
<th>((n_8, n_9))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current demand (veh/h)</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>Current travel time (min)</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

We like to route 360 freight loads from origin \((n_0)\), to destination \((n_9)\), i.e., \(D_{n_0} = 360, D_{n_1} = D_{n_2} = D_{n_3} = 0, D_{n_9} = -360\) at a minimum total cost over a time horizon of one hour. The cost in this example is defined as the travel time measured in minutes. Given the identical characteristics of the three routes it is intuitively obvious that the lowest total cost decision is to route the 360 loads equally among the three possible routes. The iterative procedure of Fig. 1 generates exactly this final decision within three iterations as shown in Table II. The total cost is considered to be the total time for all freight loads to reach their destination.

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Number of freight (Route 1)</th>
<th>Number of freight (Route 2)</th>
<th>Number of freight (Route 3)</th>
<th>Updated cost in min (Route 1)</th>
<th>Updated cost in min (Route 2)</th>
<th>Updated cost in min (Route 3)</th>
<th>Total cost (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td>2</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td>3</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>75</td>
</tr>
</tbody>
</table>

More practical example is when the three possible routes in Fig. 3 have different characteristics and traffic conditions during different time intervals as shown in Table III. The traffic simulator of Fig. 1 generates the travel times shown in Table III. The time intervals have one-hour length for a total of three consecutive hours. In this example, the demand of the current traffic exceeds the capacity for the routes at different time intervals, which indicates the presence of congestion that leads to increases in travel time. The capacity of a link is defined as the maximum number of car units per time period which can reasonably be expected to traverse a uniform section of a roadway during a given time period and it depends on many parameters such as number of lanes. In this case the distribution of the loads among the three routes is not as obvious as before. Let us assume in this case that the demand/supply is 1200 freight loads to be routed over a period of three hours i.e. \(D_{n_0} = 1200, D_{n_1} = D_{n_2} = D_{n_3} = 0, D_{n_9} = -1200\).

In this case the freight loads will significantly upset the existing travel times in a way that is not clear due to the nonlinear relationships involved between demand, capacity and travel times. The simulator however can be used to capture these nonlinear relationships and generate estimates of travel times that are far more accurate than any other analytical simplified approximation technique. The results of the iterative procedure of Fig. 1 are presented in Table IV.

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Time Interval</th>
<th>Route</th>
<th>Length (mile)</th>
<th>Capacity (veh/h)</th>
<th>Current demand (veh/h)</th>
<th>Travel time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1000</td>
<td>1200</td>
<td>1200</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>10</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>11</td>
<td>1000</td>
<td>1500</td>
<td>1500</td>
<td>31</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>12</td>
<td>1100</td>
<td>2000</td>
<td>2000</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>10</td>
<td>1000</td>
<td>950</td>
<td>950</td>
<td>17</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>11</td>
<td>1100</td>
<td>1000</td>
<td>1000</td>
<td>21</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>12</td>
<td>1100</td>
<td>960</td>
<td>960</td>
<td>17</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>10</td>
<td>1000</td>
<td>900</td>
<td>900</td>
<td>15</td>
</tr>
</tbody>
</table>

The algorithm starts with finding a solution for time interval 1, then searches for the solution in the subsequent intervals. First, route 2 is chosen since it is the minimum cost route in time interval 1. Then, 1200 units of freight are assigned to it in order to satisfy the demand. The initial choice of routing all loads using route 2 over time interval 1 led to a total travel time of 64680 minutes. Then, the link volumes are updated and route 1 happens to be the new minimum route cost from the origin to destination. Therefore, some of the freight is moved to the new route to equalize the cost of routes 1 and 2. This iterative process continues until the iteration where the total costs have been reduced to 25160 minutes.

We stopped at the 9th iteration because additional iterations led to approximately the same total cost.

### Simulation models

The simulation model framework consists of a macroscopic traffic flow and a railway simulation system. In this paper, we use the macroscopic traffic simulator (VISUM) [27] to simulate traffic flows of a road network in order to achieve fast computations and cover a wider network versus using a microscopic simulator based on VISSIM for example which may provide more accuracy but it is computationally intensive and its use is limited to small networks. The inputs to the VISUM traffic simulator are the Origin-Destination (OD) matrices, and the outputs are the estimated link volumes and corresponding travel times. The transportation network consists of several elements such as links, nodes, zones, etc. Nodes are connected by links, and links represent streets or freeways. Zones are places that a considerable number of people visit such as schools, stadiums, commercial buildings, and so on. One zone is defined for each residential district. The OD matrix determines the number of trips within zones in each time interval.

Note that vehicles’ travel times have a nonlinear relationship with the volume. We capture these nonlinearities using the simulator. The travel time of link \(l \in L\) in the loaded network \((t_{l\text{-load}})\) can be calculated as follows [28]:

\[
t_{l\text{-load}}(k + 1) = t_{l\text{-free}}[1 - q_l(k + 1)]^{q_l}\frac{1}{q_l}\max\quad (17)
\]

where \(t_{l\text{-free}}\) represents the travel time of link \(l \in L\) in the unloaded network and \(q_l(k + 1)\) is the ratio of the density at time
\( k + 1 \) to the density jam of a link i.e. \( q_{l,\text{max}} \). The parameter \( q_l \) represents the density of roadway link \( l \) which is defined as the number of vehicles per unit length. The density jam for a link occurs when the link’s volume and the link’s speed reach zero. The density jam refers to the completely stopped traffic flow scenario. The parameter \( \alpha \) is a constant parameter less than negative 1 which can be estimated using historical traffic data. Equation (17) is used to generate the travel times by using the link densities generated by the simulator. In the simulator, \( t_{l,\text{free}} \) is calculated by taking into account the length of the link and speed limit; whereas, \( q_{l,\text{max}} \) and \( \alpha \) are estimated based on the characteristic of the links such as highways or arterial streets.

We use the railway simulation system of Lu et al. [29] which is developed using the ARENA simulation package. The simulator was developed to evaluate train movements in a complex rail network. The railway track is decomposed into different segments. Then, an abstract graph is constructed where each node includes segments, junctions or stations, and each arc represents the connection between different segments or stations. Train movements in the physical railway system are simulated on movements in the constructed graph. The train travel time is calculated according to [29] by taking into account the train dynamics.

Travel times of trucks and trains are time dependent and non-linear with respect to the volumes. The outputs of the simulation framework are fed into the optimization formulation to find the best solution for the coordinated MDFLB with regard to the constraints. In our case for the two modes (i.e. rail and truck), the proposed methodology can be used in two stages. In stage one the split of the freight between the road and rail network can be performed using Fig. 1 where predicted loads and traffic conditions based on historical data and other available information are used by the simulators to generate travel times and estimated link costs to be used by the optimization procedure in an iterative manner. The final decision about the split cannot be changed in subsequent times due to practical reasons. That is, what is routed to the rail network cannot be switched back to the road network once the process starts. Within each network especially in the case of the road network incidents and unpredictable events may necessitate a stage II routing by repeating the methodology of Fig. 1 within a single network. In this more dynamic environment some changes on the fly could be possible by again evaluating the overall cost of the system. In the examples presented we assume the cost is travel time. The methodology may include costs that penalize impact on the environment, costs in meeting time windows etc.

### IV. CASE SIMULATION STUDY

The simulator framework covers an area that includes the Port complex of LA/LB, the freeway and the surface street network of about 80 square miles. We have formulated the simulation as a work-flow composition and configure the system for evaluating various scenarios for LA/LB. Fig. 4 illustrates the region of study.

We assigned 6 main destinations in the region with different demand requirements, which are supplied by three terminals in the Port complex. The first three destinations are located along I-405 highway between the I-110 and I-710 freeways; whereas, the other three destinations are placed along CA-91 as shown in Fig. 4 by D. The simulation framework consists of VISUM and the railway simulation system is used to estimate the link travel time in the region of study. We assume that 5 trains with homogenous capacities of 50 containers each are available at the port complex. Average weight of the containers is assumed to be 25 tonnes. The transportation cost per unit (price/(ton.mile)) for truck transportation is estimated to be 37 cents; whereas, train transportation is much cheaper to about 3 cents [30]. The requirements of the destinations are provided in Table V in terms of containers.

<table>
<thead>
<tr>
<th>Dest. 1</th>
<th>Dest. 2</th>
<th>Dest. 3</th>
<th>Dest. 4</th>
<th>Dest. 5</th>
<th>Dest. 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>350</td>
<td>450</td>
<td>400</td>
<td>600</td>
<td>700</td>
<td>560</td>
</tr>
</tbody>
</table>

Let the three terminals (A, B, C) be the origins and none of the shipping companies (SC) have any knowledge of the other’s scheduling and routings decisions. We set one coordinator that collects and updates the information from the shipping companies for the coordinated MDFLB based on the origins-and-destinations of freight as shown in Fig. 4. The total demand requirement is 3060 containers which are equally distributed among the origins (e.g. each origin has 1020 containers). Links \( x_1 \) and \( x_2 \) in Fig. 5, correspond to the I-710 and I-110 freeways respectively. A set of links connects a pair of origin-destination in the constructed network. Two transportation modes are taken into account for the experiments in this paper e.g. trucks and trains. We model the double-track railway system as the constructed railway network as shown in Fig. 6. Containers are transferred to the destinations by two different ways. Either they are delivered directly by trucks from the origins to the destinations or they are transferred partially by trains as illustrated in Fig. 6. The solid line indicates the aggregated roads and the dotted line represents a railway in the constructed network.
Accidents are introduced in the region of study for volumes from the southern California Association of Governments (SCAG). Traffic data is only available for a small portion of the network are assigned in VISUM as the baseline for the scenarios which includes rush hour, noon and night conditions.

Five scenarios are defined for different situations within a day which have time dependent traffic conditions in the region. Rush hour, noon, and night-time correspond to the first three scenarios respectively. Accidents are introduced in the region of study for the fourth scenario, and real time traffic volume for an individual link changes significantly during the optimization procedure in the fifth scenario. The system updates solutions every hour to capture new conditions.

Initially, traffic volumes for all the links in the transportation network are assigned in VISUM as the baseline for the scenarios obtained from the Southern California Association of Governments (SCAG). Traffic data is only available for a small portion of the links in the selected region; hence, dynamic traffic assignment is used to estimate volumes for the other links. Then, trucks/trains are loaded into the network dynamically using the optimization algorithm and the integrated simulators. First, we assume that interaction exists among the shipping companies, and each of these companies operates independently (uncoordinated). Then, the coordinator collects the origins-to-destinations information from all the shipping companies and determines the MDFLB (coordinated). Table VI illustrates the role of the coordinator that coordinates the shipping companies’ data. Traffic condition is set to be rush hour.

**Table VI**

<table>
<thead>
<tr>
<th>Destination</th>
<th>Uncoordinated</th>
<th>Coordinated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Final sol. Travel time (min)</td>
<td>Number of possible routes chosen to each destination</td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>28</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>41</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>42</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>41</td>
<td>17</td>
</tr>
</tbody>
</table>

The first column represents destination number. The second column gives the final solution in terms of average travel time of transferring containers from the origins to the assigned destination using all the transportation modes (e.g. trucks and trains) without incorporating the coordinator. The third and sixth columns are the number of possible routes chosen to each destination. The fifth column represents average travel time by taking into account information sharing between the shipping companies and the coordinator. To verify the results from the optimization models, we use a simulated system which consists of the microscopic traffic flow simulator (VISSIM) and the railway simulation system as representation of the real system shown in Fig. 1 [27]. The final solution is what is obtained by the microscopic simulator which in the framework of Fig. 1 serves as the actual system and the simulated travel time is the one generated by the simulation model.

The microscopic traffic flow simulator (VISSIM) simulates single vehicle-driver unit which takes into account the dynamics of the individual vehicles. In VISSIM, each entity e.g. car, truck, bus, person is simulated individually. The fourth and seventh columns of Table VI indicate the average travel time of the uncoordinated and coordinated MDFLB using the simulated system respectively. In the optimization models, the travel times are estimated using the macroscopic traffic flow simulator and the railway simulation system. Our methodology uses the macroscopic traffic flow simulator since the microscopic traffic flow simulator would be computationally intensive to use in our iterative approach. To test the accuracy of these estimates, we take the final volumes from the optimization solution and feed them to the simulated system to estimate the travel times of the actual system. As Table VI shows, the results from columns 2 and 4 and columns 5 and 7 closely match each other suggesting that using a macroscopic traffic simulator is sufficient to estimate the parameters of the optimization model. Fig. 7 demonstrates the convergence of the iterative approach for the coordinated MDFLB in the rush hour condition. The total cost (transportation cost) is reduced by more than 49% from the initial solution using the proposed methodology. A change in the total cost less than $1000 between two consecutive iterations is considered as the stopping criterion in this case.
As mentioned earlier, the links’ volume in the network are time-dependent and change due to various reasons such as load deployment and incidents. To demonstrate the time varying characteristic of the links’ volume, two links are selected as shown in Fig. 7 ($x_1, x_2$). Links $x_1$ and $x_2$ correspond to the I-710 and B2110 freeways respectively. Containers are deployed within the time interval $k = [0, 5]$. Fig. 8 illustrates the links’ volumes with respect to time for the uncoordinated and coordinated MDFLB in the rush hour scenario. Note that in the coordinated MDFLB, the link volumes $x_1$ and $x_2$ are more balanced between them than in the uncoordinated case where link volume $x_1$ varies greatly.

![Uncoordinated](image1.png)

![Coordinated](image2.png)

Fig. 8. (a) Uncoordinated MDFLB (b) Coordinated MDFLB

The following tables demonstrate the average travel times of transferring containers from the origins to the assigned destinations using trucks and trains for each scenario which corresponds to different time-periods within a day (rush hour, noon, night time).

<table>
<thead>
<tr>
<th>TABLE VII</th>
<th>RUSH HOUR</th>
<th>Destination</th>
<th>Initial solution Travel time (min), no opt.</th>
<th>Number of routes chosen to each destination, no opt.</th>
<th>Final solution Travel time (min), opt.</th>
<th>Number of routes chosen to each destination, opt.</th>
<th>Percentage improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>49</td>
<td>3</td>
<td>23</td>
<td>19</td>
<td>3</td>
<td>19</td>
<td>64</td>
</tr>
<tr>
<td>2</td>
<td>41</td>
<td>3</td>
<td>22</td>
<td>17</td>
<td>4</td>
<td>17</td>
<td>63</td>
</tr>
<tr>
<td>3</td>
<td>38</td>
<td>3</td>
<td>21</td>
<td>16</td>
<td>5</td>
<td>16</td>
<td>62</td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>3</td>
<td>31</td>
<td>24</td>
<td>6</td>
<td>24</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>58</td>
<td>3</td>
<td>29</td>
<td>26</td>
<td>7</td>
<td>26</td>
<td>58</td>
</tr>
<tr>
<td>6</td>
<td>57</td>
<td>3</td>
<td>30</td>
<td>27</td>
<td>8</td>
<td>27</td>
<td>55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE VIII</th>
<th>NOON</th>
<th>Destination</th>
<th>Initial solution Travel time (min), no opt.</th>
<th>Number of routes chosen to each destination, no opt.</th>
<th>Final solution Travel time (min), opt.</th>
<th>Number of routes chosen to each destination, opt.</th>
<th>Percentage improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24</td>
<td>3</td>
<td>13</td>
<td>15</td>
<td>46</td>
<td>15</td>
<td>64</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>3</td>
<td>13</td>
<td>14</td>
<td>44</td>
<td>14</td>
<td>61</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>3</td>
<td>12</td>
<td>12</td>
<td>40</td>
<td>12</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>26</td>
<td>3</td>
<td>18</td>
<td>20</td>
<td>31</td>
<td>20</td>
<td>56</td>
</tr>
<tr>
<td>5</td>
<td>27</td>
<td>3</td>
<td>17</td>
<td>20</td>
<td>37</td>
<td>20</td>
<td>55</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>3</td>
<td>16</td>
<td>18</td>
<td>36</td>
<td>18</td>
<td>50</td>
</tr>
</tbody>
</table>

The first column represents destination number which includes 6 main destinations in the region as shown in Fig. 4. The second column indicates the initial solution (average travel time) in terms of minutes when no optimization is applied for the containers’ transfer decision. For the initial solution, the fastest route to each destination is found for a single container. Then, all the containers are transferred to their destinations using that route i.e. without taking into account the impact of the load on travel time and shortest route. The third column represents the number of routes chosen to each destination with no optimization taken into account. As stated earlier, only one route (the fastest route) is chosen to deliver demands from each origin to each destination. Two transportation modes (train and truck) are taken into consideration for transferring containers, and three types of nodes (e.g. origins, destinations, intermediary) are defined in the network where train stations are in the intermediary node category. The fourth column indicates the final solution (average travel time) in terms of minutes from the optimization models to transfer all the demand from the origins to the assigned destinations using both transportation modes. The fifth column is the number of routes chosen to each destination using the optimization models and the integrated simulators. For instance, 19 different routes are chosen to deliver containers from the port complex to destination 1 in the rush hour scenario. The number of routes and volumes which are assigned to each route are obtained from the final solution of the optimization model. Finally, the last column indicates the percentage improvement achieved by the final solution over the initial solution.

In the fourth scenario, it is assumed that accidents occur during the rush hour. Accidents are introduced on two main freeways (I-110 and 1710) and a main street causing the capacity of these links to reduce by a half during the rush hour period. The locations of the accidents are shown in Fig. 9 by circles.

Table X demonstrates the results for the fourth scenario (accidents during the rush hour period). In this case, the coordinated MDFLB finds solutions to cope with the accidents. As expected, the percentage improvement between the initial and final solutions increases in situations with traffic congestion in some links due to rush hour and accidents. The same approach can be applied for other events such as road closure.
In this paper we developed a Multimodal Dynamic Freight Load Balancing (MDFLB) system to balance freight loads across the rail and road network. In the coordinated MDFLB, a coordinator collects and updates information from all the shipping companies to make transfer decisions. We use optimization and network simulators in an iterative manner to find the best coordinated solution to transfer freight from a set of origins to a set of destinations. What makes the optimization problem more complex is that the cost function itself depends on the states of the network that vary in time and space and are affected by load balancing decisions. A balancing decision based on a measured or predicted state of the network may change the total load on certain links, lead to new states and therefore new cost variables. This change necessitates exercising the optimization again hence the iterative procedure. We developed the approach for two networks rail and road network. We use macroscopic traffic simulators for the iterative procedure and a microscopic traffic simulator as the actual system for testing purposes. The same rail network simulator was used for testing and final solution. The proposed load balancing system is demonstrated using an example of load distribution in the Los Angeles/Long Beach area that involves the two major Ports.

The computational results demonstrated the effectiveness of the proposed approach in reducing traffic congestion and average travel times.

In this paper we did not assume any strict time windows for load pick up/delivery. The problem with time windows will add to complexity but it will not change the proposed concept of dynamic load balancing. The added complexity will give rise to scaling and computational issues which are currently under investigation.

### References


Afshin Abadi is a Ph.D. student at the University of Southern California. He received his M.S. in Electrical Engineering from the University of Southern California. His current research focuses on container terminal operation, transportation planning, and car navigation system.