

# **Optimum Routing of Freight in Urban Environments under Normal Operations and Disruptions using a Co-simulation Optimization Control Approach**

FINAL REPORT

METRANS Project DTRT13-G-UTC57

By

Petros Ioannou (Principal Investigator)

University of Southern California

Electrical Engineering – Systems, EEB 200B

Los Angeles, CA 900089-2562

December 27, 2016



## **Disclaimer**

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, and California Department of Transportation in the interest of information exchange. The U.S. Government and California Department of Transportation assume no liability for the contents or use thereof. The contents do not necessarily reflect the official views or policies of the State of California or the Department of Transportation. This report does not constitute a standard, specification, or regulation.

## Abstract

The complexity and dynamics of multimodal freight transportation together with the unpredictability of incidents, disruptions and demand changes make the optimum routing of freight a challenging task. Optimum routing decisions in a multimodal transportation rely on the estimation of the dynamical states of the multimodal traffic network. Such estimations rely on mathematical models that are often highly inaccurate leading to routing decisions that deviate considerably from optimality. In addition, they do not consider the impact of the routed freight on the states of the network.

The purpose of this project is to use complex real time simulation models to estimate the states of the transportation network and integrate that information with optimization and load balancing techniques in an iterative feedback configuration that would lead to much more efficient routing decisions during normal operations and disruptions. The approach is referred to as the CO-SiMulation Optimization (COSMO) approach. We use a simulation testbed consisting of a road traffic simulation model and a rail simulation model for the Los Angeles/Long Beach Port area to demonstrate the efficiency of the proposed approach. The results demonstrate the potential of the approach for practical freight routing.

## Table of Contents

Disclaimer .....	II
Abstract .....	III
List of Tables .....	V
List of Figures.....	VI
Disclosure .....	VII
Acknowledgments .....	VIII
1 Introduction.....	1
2 Problem Formulation .....	4
3 Proposed COSMO Approach .....	8
3.1 The Overview of COSMO Approach .....	8
3.2 Find Minimum Cost Route .....	11
3.3 Load Balancing Algorithm .....	14
4 Evaluation Results .....	18
4.1 Simulation Models .....	18
4.2 Case Study and Results.....	19
5 Conclusion and Summary.....	22
6 Implementation.....	23
References .....	24

## List of Tables

Table 1:	Route Characteristics and Traffic Conditions of Simple Example .....	15
Table 2:	Demands of Destinations .....	20
Table 3:	Simulation Results of Three Cases.....	21

## List of Figures

Figure 1: COSMO iterative approach for optimum multimodal routing .....	8
Figure 2: Performances of different load balancing algorithms.....	17
Figure 3: Road network simulation .....	18
Figure 4: Region of study.....	20

## Disclosure

Project entitled ‘Optimum Routing of Freight in Urban Environments’ was funded in entirety under a contract to California Department of Transportation. The PI of the project was Professor Petros Ioannou of the Viterbi School of Engineering at the University of Southern California and was carried out during the period : 07/01/2015 - 06/30/2016. The total amount of funding of the project was \$100,000.

## Acknowledgments

We would like to thank METRANS for funding this project and acknowledge the Southern California Association of Governments (SCAG) for providing the historical road traffic data. We are also like to acknowledge the help of the PhD students Yanbo Zhao and Yihang Zhang and students Liang Liu and Santiago Carvajal who assisted with the data analysis and simulations.



# 1 Introduction

Efficient freight movement is an essential factor not only in urban transportation but also in social and economic development as well as environmental considerations [1-3]. The growth of worldwide trade will significantly increase traffic congestion and air pollution especially in metropolitan areas with major ports such as Los Angeles/Long Beach where there is a high concentration of both freight and passenger traffic that share the same infrastructure. One of the biggest challenges for freight transport efficiency in such multimodal environments is the fact that the same rail and road networks are used for moving people and goods which leads to non-homogeneous traffic. This non-homogeneity has a detrimental impact on the transportation system performance because of the differences of vehicle sizes and dynamics between passenger and freight transport. The freight vehicles such as freight trains and trucks take longer distances to stop, longer time to accelerate from a stopping position, consume more fuel and generate more air pollution compared to passenger vehicles. The situation becomes even worse during incidents and disruptions that lead to network changes such as road or railway closures that require rapid response and distribution of freight traffic across the multimodal network. Without efficient routing of freight, the transportation network will face severe capacity shortages, inefficiencies, and route load imbalances across the network in space and time. A more efficient freight routing system could reduce transport costs and contribute to sustainability and efficiency of the entire urban transportation network.

Due to the important role of freight transportation, numerous researchers have addressed the issue of multimodal freight routing and scheduling [1-24]. Jourquin and Beuthe presented a multimodal freight model based on a digitized geographic network [1]. Southworth and Peterson developed a multi-layer intermodal shipment routing model in [2]. The intermodal freight transport between rail and road has been described in [4-6]. As a fundamental issue for optimum routing, the problem of finding the shortest path of an origin-destination pair in a multimodal network has been studied by many researchers. Modesti and Sciomachen applied a link utility measure approach to solve the multi-objective shortest path problem [7]. Lozano and Storchi considered the impact of modal transfer costs when finding the shortest multimodal path [8]. Dynamic and stochastic routing for a multimodal transportation environment was studied in [9] and [10].

Speed-up techniques for finding the shortest path have also been analyzed including Core-Based routing [11], label-setting and label-correcting methods [12] and the improved label setting algorithms [13]. Optimization techniques have been commonly used to solve the multimodal transport routing and scheduling problem such as in [14-17]. Castelli et al. used a Lagrangian-based heuristic procedure to solve the signal line and general network scheduling problem [14]. Zografos et al. developed a dynamic programming based algorithm for multimodal time-scheduling with a shortest path algorithm [16]. Moccia et al. solved a multimodal routing problem with timetables and time windows by integrating heuristics and a column generation algorithm [17]. The main difficulty in these research efforts and past work is that the classical approach of using mathematical models breaks down when faced with the control and optimization of complex networks such as multimodal freight networks that exhibit nonlinear travel time functions which are difficult to mathematically represent and to find a closed form solution. Some researchers tried to solve the multimodal scheduling problem from the aspect of user equilibrium conditions with logit-based or probit-based choice models as in the unimodal road scheduling [18-22]. Moreover, Russ et al. and Yamada et al. showed the applications of routing and scheduling in multimodal freight network design [23-24].

The availability of fast computers and software tools opens the way for new approaches which can overcome the today's limitations of network modeling complexity. The traffic flows and states can be better predicted using simulation models that are more complex and can capture phenomena that cannot be formulated with simple models [25]. These simulation models can also be integrated with control and optimization techniques to provide better and more robust decisions in a real time manner. The purpose of this project is to formulate and solve the dynamical multimodal freight routing problem by exploiting the availability of powerful computational software tools. We therefore propose a method we refer to as COSMO (CO-Simulation Optimization) as a potential innovation in dealing with multimodal transportation routing that cannot be handled by the traditional way. We are proposing a novel load balancing algorithm in the COSMO approach which uses the real time simulation models to estimate the marginal costs of the routes which are used by an optimization algorithm to calculate the minimum cost route. The load balancing algorithm, real time simulation model and optimization module are connected in a feedback iterative loop where the load distribution in the network is modified at each iteration in order to reduce the overall cost. The process is adaptive as in the case of incidents or

unexpected demand changes or changes in infrastructure due to emergencies sudden or planned the simulation model captures the impact on the network and re-evaluates its states and cost estimations that feed the optimization module.

This report is organized as follow. Section 2 gives a formulation of the optimum routing problem for multimodal freight transport. Section 3 proposes the COSMO approach and demonstrates how to solve the formulated problem with the proposed approach. Section 4 shows the experimental results of the proposed approach for a scenario in Southern California. Finally, the conclusions are discussed in Section 5.

## 2 Problem Formulation

A multimodal freight transportation network  $G$  can be represented as a directed graph network consisting of a set of nodes ( $N$ ) with a set of directed links ( $L$ ) connecting the nodes. A link in the network could be one segment of a roadway or railway track while a node with zero length connects multiple links. All passenger and freight traffic start and end at certain network nodes. Let  $I$  and  $J$  be the sets of origin and destination nodes respectively. Both  $I$  and  $J$  are a subset of node set  $N$ . In this report we are dealing with the routing of freight traffic that are container flows between origin and destination nodes. The analysis time horizon is discretized into  $|K|$  small time intervals. The notation that will be used throughout the report is defined as follows:

- $i$  The index of an origin node,  $i \in I$ ;
- $j$  The index of a destination node,  $j \in J$ ;
- $k$  The index of time,  $k \in K$  where  $K = \{0, 1, \dots, |K|\}$ ;
- $l$  The index of a link in the network,  $l \in L$ ;
- $P_{i,j}$  The set of all paths from an origin  $i$  to a destination  $j$ ;
- $p$  The index of a path from an origin  $i$  to a destination  $j$ ,  $p \in P_{i,j}$ ;
- $\Delta t$  The length of a time interval (unit: hour);
- $d_{i,j}$  The total number of containers departing from origin node  $i$  to destination node  $j$ ;
- $x_{i,j}^p(k)$  The number of containers departing from origin node  $i$  to destination node  $j$  using path  $p$  with a departure time of  $k$ ;
- $y_l(k)$  The traffic volume of link  $l$  at time  $k$ ;
- $w_l(k)$  The travel time of link  $l$  at time  $k$ ;

We next describe the constraints for the formulated multimodal freight transport problem. The first constraint ensures that the total amount shipped throughout the day for each origin/destination pair equals the required number.

$$\sum_k \sum_{p \in P_{i,j}} x_{i,j}^p(k) = d_{i,j}, \quad \text{for } \forall i \in I, \forall j \in J \quad (1)$$

Let  $Y(k) = [y_1(k), y_2(k), \dots, y_{|L|}(k)]'$  be the vector of traffic volumes on links 1 to  $|L|$  at time  $k$ .

Then the relationship of the traffic volume on a link  $l$  with the departure container traffic and other parameters in the network can be expressed as a nonlinear dynamical equation:

$$y_l(k+1) = f_l(y_l(k), a_l(k), X(k), k), \text{ for } \forall l \in L, \forall k \in K \quad (2)$$

where

$$X(k) = [x_{i,j}^p(k) : \forall i \in I, \forall j \in J, \forall p \in P_{i,j}]' \quad (3)$$

In (2),  $f_l$  is a nonlinear and time-dependent function of the traffic volume of a link  $l \in L$ . The impact of the traffic volumes from adjacent links at time  $k$  is denoted by  $a_l(k)$  and  $X(k)$  is the vector of departure traffic volumes from all origin nodes at time  $k$  as in (3). Since  $y_l(k)$  and  $a_l(k)$  contain the impact of the previous departure container traffic before time  $k$  (i.e.,  $X(r)$  for  $\forall r < k$ ) so only  $X(k)$  is included in equation (2). The link volumes in the transportation network are time-dependent due to various factors such as traffic network changes, accidents and incidents.

Let  $W(k) = [w_1(k), w_2(k), \dots, w_{|L|}(k)]'$  be the vector of travel time (unit:  $\Delta t$ ) of links 1 to  $|L|$  at time  $k$ .  $w_l(k)$  specifies the time length that a container takes to travel on link  $l$  if it enters link  $l$  at time  $k$ . The link travel time is a function of the link volume at time  $k$  which is time-dependent because of the impact of the time-dependent passenger traffic, network incidents and railway dispatching decisions.

$$W(k) = g(Y(k), k), \text{ for } \forall k \in K \quad (4)$$

Let  $T_{i,j}^p(k)$  be the travel time (unit:  $\Delta t$ ) of a path  $p$  from an origin node  $i$  to destination node  $j$  if a container departs origin  $i$  at time  $k$ . Assume a path  $p$  contains links  $l_{p,1} \rightarrow \dots \rightarrow l_{p,N_p}$  where  $N_p$  is the number of links on this path  $p$ . Define  $e_{l_{p,n_p}}(k)$  as the entering time at link  $l_{p,n_p}$  for a container on path  $p$  with a departure time of  $k$  at the origin. Then the path travel time can be computed as follows:

$$T_{i,j}^p(k) = \sum_{n_p=1}^{N_p} w_{l_{p,n_p}} \left( e_{l_{p,n_p}}(k) \right) \quad (5)$$

where

$$e_{l_{p,1}}(k) = k, \text{ and } e_{l_{p,n_p+1}}(k) = e_{l_{p,n_p}}(k) + w_{l_{p,n_p}} \left( e_{l_{p,n_p}}(k) \right), \text{ for } n_p = 1, \dots, N_p - 1 \quad (6)$$

Let  $S_{i,j}^p(k)$  be the total cost for moving one container from origin  $i$  to destination  $j$  with a departure time of  $k$ . Then the objective function can be expressed as

$$\sum_{k \in K} \sum_{i \in I} \sum_{j \in J} \sum_{p \in P_{i,j}} S_{i,j}^p(k) x_{i,j}^p(k) \quad (7)$$

where

$$S_{i,j}^p(k) = C_{i,j}^p(k) + K_p T_{i,j}^p(k), \text{ for } \forall i \in I, \forall j \in J, \forall p \in P_{i,j}, \forall k \in K \quad (8)$$

In (8),  $C_{i,j}^p(k)$  is the non-travel time cost (unit: dollar) generated by vehicle usage, distance cost, etc. per container departing from origin  $i$  to destination  $j$  at time  $k$  using path  $p$ . This cost could be obtained directly based on the path distance and used vehicle type.  $K_p$  is the weight value of the travel time on path  $p$ .

In summary, the multimodal routing problem can be expressed as follows.

$$\min \sum_{k \in K} \sum_{i \in I} \sum_{j \in J} \sum_{p \in P_{i,j}} S_{i,j}^p(k) x_{i,j}^p(k) \quad (9)$$

subject to constraints (1) – (6) and

$$x_{i,j}^p(k) \geq 0 \text{ for } \forall i \in I, \forall j \in J, \forall p \in P_{i,j}, \forall k \in K \quad (10)$$

$$0 \leq y_l(k) \leq u_l(k) \text{ for } \forall l \in L, \forall k \in K \quad (11)$$

$$\text{given } d_{i,j}, u_l(k) \text{ for } \forall i \in I, \forall j \in J, \forall l \in L, \forall k \in K \quad (12)$$

where  $u_l(k)$  is the capacity of link  $l$  at time  $k$ .

The explicit forms of the dynamical functions in (2) and (4) are difficult to mathematically express directly due to the nonlinearities and complex variable interactions. Therefore, we propose a COSMO approach in which the traffic network simulation models are used to replace

the mathematical functions of (2) and (4) to generate more accurate link volumes and predict link/path travel times to solve the above optimization problem iteratively.

.

### 3 Proposed COSMO Approach

#### 3.1 The Overview of COSMO Approach

Figure 1 describes our approach. Real time data that reflect changes in demand or incidents or traffic changes are fed to the simulation model which updates itself continuously. The simulation model generates the estimate flows in all links of the network under consideration and uses the estimated states to calculate the costs along every link. This information is used by the optimization to calculate the minimum cost route for an initial choice of freight load distribution. Since the freight load will affect the link cost in a nonlinear manner that is not known the initial choice of a minimum cost route may no longer be minimum after the deployment of freight load. If this is the case the load balancing module makes an adjustment which leads to new link costs and therefore to possible new routes of minimum cost as computed by the optimization part. This leads to an iterative procedure which continues till certain stopping criteria are satisfied. The final solution then becomes ready to be applied to the actual freight network. Unpredictable changes in the freight network characteristics are captured by the model and taken into account therefore the proposed approach can deal with disruptions and all kind of emergencies and incidents in the network.

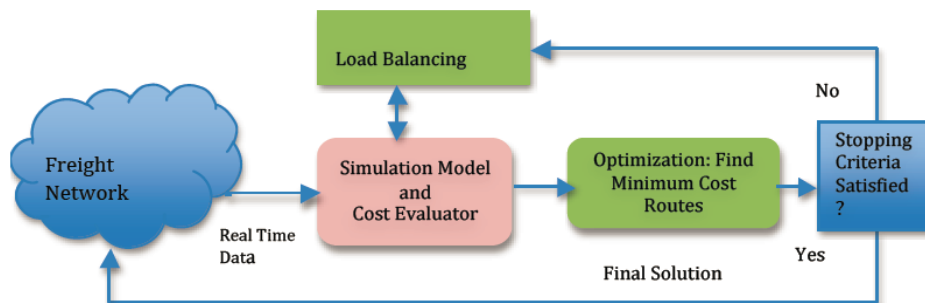


Figure 1: COSMO iterative approach for optimum multimodal routing

Below we present the mathematical description of the various modules in the COSMO approach as applied to freight routing.



Let  $\varphi_{i,j,k}^p(X)$  be the predicted average cost per container of  $x_{i,j}^p(k)$  in a routing decision  $X$  from the updated link states based on the simulation output where  $X$  is a routing decision from all the origins to the destinations for the entire analysis horizon, i.e.

$$X = \left[ x_{i,j}^p(k), \forall i \in I, j \in J, p \in P_{i,j}, k \in K \right] \quad (13)$$

Since the network simulation models can predict the generated link volumes and travel times under the constraints of the traffic flow dynamics (2) - (6) and link capacities (11), the original optimization problem can be rewritten as:

$$\min TC(X) = \sum_{k \in K} \sum_{i \in I} \sum_{j \in J} \sum_{p \in P_{i,j}} \varphi_{i,j,k}^p(X) x_{i,j}^p(k) \quad (14)$$

subject to

$$\sum_k \sum_{p \in P_{i,j}} x_{i,j}^p(k) = d_{i,j}, \quad \text{for } \forall i \in I, \forall j \in J \quad (15)$$

$$x_{i,j}^p(k) \geq 0, \quad \text{for } \forall i \in I, \forall j \in J, \forall p \in P_{i,j}, \forall k \in K \quad (16)$$

$$\text{given } d_{i,j} \quad \text{for } \forall i \in I, \forall j \in J \quad (17)$$

Assume that the cost function is continuous and differentiable, the first-order necessary conditions for an optimal solution  $X^*$  of the above problem are:

$$\sum_{k' \in K} \sum_{i' \in I} \sum_{j' \in J} \sum_{p' \in P_{i',j'}} \left. \frac{\partial TC(X)}{\partial x_{i',j'}^{p'}(k')} \right|_{X=X^*} \left( x_{i',j'}^{p'}(k') - x_{i',j'}^{p'}(k')^* \right) \geq 0, \quad (18)$$

$$\text{for } \forall x_{i',j'}^{p'}(k') \in \Omega$$

where  $\Omega$  is the feasible solution set given by constraints (15)-(17). The first order derivative in (18) gives the change in the total cost by adding one more container on path  $p'$  with a departure time of  $k'$  from origin  $i'$  to destination  $j'$ , i.e. the marginal route cost of an OD pair. The conditions (18) mean that the total cost cannot be reduced further by switching the optimal solution  $X^*$  to another solution locally.

The conditions (18) also state that at the optimal solution, the marginal route costs of any used paths connecting the same OD pair are equal and less than or equal to the marginal total cost of any other unused paths. Otherwise, there exists a neighborhood solution such that the total cost can be reduced further. In other words, by redistributing some containers from the used routes having greater marginal route costs to other routes having smaller marginal path costs, the total cost may be reduced. The idea of the proposed COSMO approach is to find the routes with minimum marginal costs and then conduct the load balancing from the current used routes to the new found routes with less marginal path costs in order to reduce the total cost. The overall steps of the COSMO approach can be described as follows:

Step 1: Obtain an initial solution

Set the iteration counter  $m = 0$ . Assign the given freight flows to a subset of predefined routes in the transportation network, and obtain an initial routing decision  $X^{(0)}$ . As an example, the predefined route for each OD pair can be the route with minimum cost before adding the freight traffic.  $P_{i,j}(k)^{(0)}$  is the set of the used routes in the initial decision at time  $k$ .

Step 2: Update the link states

Set the current routing decision  $X^{(m)}$  into the freight network simulation models and run the simulation models to obtain updated link volumes, travel times as well as other link states.

Step 3: Search for new minimum route

With the updated link costs, find new time-dependent routes with a smaller marginal cost. Let  $p_r$  be a new found route for an OD pair  $(i, j)$  and its departure time be  $k_r$ .

$$P_{i,j}(k)^{(m+1)} = P_{i,j}(k)^{(m)} \cup p_r \text{ if } k_r = k \quad (19)$$

Step 4: Check for convergence

Check whether the convergence criteria is satisfied, stop the algorithm if the cost reduction between two consecutive iterations is less than a predefined threshold value or the maximum number of iterations has been achieved. Otherwise, go to the load balancing step 5.

Step 5: Perform load balancing

For the OD pair with the new minimum route, redistribute the freight loads among the

current used routes connecting this OD pair to the new route with the smaller marginal cost.

Considering the fact that it is difficult to find the explicit functional form of  $TC(X)$ , the marginal costs of the different routes cannot be computed directly so they need to be estimated with the updated link states from the simulation models. We next describe how to find the new routes with the estimates of the marginal route cost from the simulation results and then provide the details of the load balancing algorithm in the following sections.

### 3.2 Find Minimum Cost Route

Assume that we have a current routing decision  $X$  and its corresponding  $\varphi_{i,j,k}^p(X)$  can be obtained from running the simulation models. Then, the marginal costs of the different routes can be computed by the following equation:

$$\begin{aligned} \frac{\partial TC(X)}{\partial x_{i',j'}^{p'}(k')} &= \frac{\partial \sum_{k \in K} \sum_{i \in I} \sum_{j \in J} \sum_{p \in P_{i,j}} \varphi_{i,j,k}^p(X) x_{i,j}^p(k)}{\partial x_{i',j'}^{p'}(k')} \\ &= C_{i',j'}^{p'}(k') + K_p T_{i',j'}^{p'}(k') + \sum_{k \in K} \sum_{i \in I} \sum_{j \in J} \sum_{p \in P_{i,j}} K_p x_{i,j}^p(k) \frac{\partial T_{i,j}^p(k)}{\partial x_{i',j'}^{p'}(k')} \end{aligned} \quad (20)$$

Equation (20) shows the change in the total cost if  $x_{i',j'}^{p'}(k')$  is changed by one container. The first term in (20) is the non-travel time cost from the vehicle usage, distance cost, etc. which is available directly and the second term is the travel time of the route that can be computed using equations (5) and (6) with the predicted link travel times from the simulation models. Thus, the values of these two terms are obtained using the simulation models. The third term describes the change in the total travel time when changing one container from origin  $i'$  to  $j'$  using path  $p'$  with a departure time of  $k'$ , which is difficult to mathematically express directly since the travel time of a given route is a complicated function of the traffic volumes of this route and other factors due to the nonlinear dynamical characteristics and route interactions of the traffic network. As a possible approach, the values of the third term in (20) can also be obtained using the simulation models, i.e. assigning one more container on this route then observing changes in the travel times of the other routes. However, it is very time consuming to obtain the impact of all the possible routes by enumerating simulations for all the routes when the network is large scale. In order to speed up the search of the new minimum routes, we propose an approximation method to

estimate the third term in (20) without running simulation models repeatedly.

When running the simulation of the current routing decisions, we obtain the current link states  $y_l(k)$ ,  $w_l(k)$ , and path travel time  $T_{i,j}^p(k)$  for  $\forall i \in I, \forall j \in J, \forall p \in P_{i,j}, \forall k \in K$ . By the derivative chain rule and equations (4) - (6),

$$\begin{aligned} \frac{\partial T_{i,j}^p(k)}{\partial x_{i',j'}^{p'}(k')} &= \sum_{n_p=1}^{N_p} \frac{\partial w_{l_{p,n_p}}(e_{l_{p,n_p}}(k))}{\partial x_{i',j'}^{p'}(k')} \\ &= \sum_{n_p=1}^{N_p} \frac{\partial w_{l_{p,n_p}}(e_{l_{p,n_p}}(k))}{\partial y_{l_{p,n_p}}} \frac{\partial y_{l_{p,n_p}}(e_{l_{p,n_p}}(k))}{\partial x_{i',j'}^{p'}(k')} \\ &\text{for } \forall i' \in I, \forall j' \in J, \forall p' \in P_{i',j'}, \forall k' \in K \end{aligned} \quad (21)$$

The term  $\frac{\partial w_l}{\partial y_l}$  in (21) is the derivative of the link travel time with respect to the link volume. It can be approximately determined using the simulated link traffic volumes  $y_l(k)$  and the link capacities  $u_l(k)$ . The calculations of the derivative of the link travel time term for road links and railway links are different due to their different characteristics. .

For the road links, the travel time derivative can be obtained using the fundamental diagram of traffic flow [26] with the observed link volume and average travel time. The travel time derivative can also be determined using a road travel time model such as the Bureau of Public Roads (BPR) function in [27] or other estimated functions in [28]. Take the BPR function as an example,

$$w_l = w_{l\_free} \times \left( 1 + \alpha \left( \frac{y_l}{u_l} \right)^\beta \right) \quad (22)$$

where  $w_l$  is the link travel time,  $w_{l\_free}$  is the link free-flow travel time that is determined by the link length and speed limit,  $y_l$  is the link volume and  $u_l$  is the link capacity.  $\alpha \geq 0, \beta \geq 0$  are model parameters that can be estimated from historical data. With this function, the link travel time derivative in (21) for a road link can be computed by the following equation:

$$\frac{\partial w_l}{\partial y_l}(k) = \left[ \frac{\alpha \beta w_{l\_free} \left( \frac{y_l(k)}{u_l(k)} \right)^{\beta-1}}{u_l(k)} \right] \quad \text{for } \forall l \in L, \forall k \in K \quad (23)$$

For the railway links, considering the impact of the passenger train schedule and the freight dispatching decisions, the travel time of a link is not an explicit function of link volumes so the corresponding travel time derivative cannot be estimated easily. Therefore the travel time derivatives of the rail links are estimated from running the rail simulation models repeatedly or using historical operational data.

The term  $\frac{\partial y_{l_{p,n_p}}(e_{l_{p,n_p}}(k))}{\partial x_{i',j'}^{p'}(k')}$  in (21) describes the derivative of the traffic volume of link  $l_{p,n_p}$  at time  $e_{l_{p,n_p}}(k)$  when  $x_{i',j'}^{p'}(k')$  changes by one unit of container. Ignoring the link interactions, we can estimate this term using the following equation:

$$\frac{\partial y_{l_{p,n_p}}(e_{l_{p,n_p}}(k))}{\partial x_{i',j'}^{p'}(k')} \approx \begin{cases} \frac{1}{\Delta t}, & l_{p,n_p} = l_{p',n_{p'}} \text{ and } e_{l_{p,n_p}}(k) = e_{l_{p',n_{p'}}}(k') \\ 0, & \text{otherwise} \end{cases} \quad (24)$$

for  $\forall i' \in I, \forall j' \in J, \forall p' \in P_{i',j'}, \forall k' \in K$

Finally, the marginal costs of a route by (20)-(24) can be approximately computed by,

$$\frac{\partial TC(X)}{\partial x_{i',j'}^{p'}(k')} \approx \sum_{n_{p'}=1}^{N_{p'}} \left( c_{l_{p',n_{p'}}}(e_{l_{p',n_{p'}}}(k')) + \kappa_{l_{p',n_{p'}}} w_{l_{p',n_{p'}}}(e_{l_{p',n_{p'}}}(k')) + \kappa_{l_{p',n_{p'}}} \frac{z_{l_{p',n_{p'}}}(e_{l_{p',n_{p'}}}(k'))}{\Delta t} \frac{\partial w_{l_{p',n_{p'}}}(e_{l_{p',n_{p'}}}(k'))}{\partial y_{l_{p',n_{p'}}}(e_{l_{p',n_{p'}}}(k'))} \right) \quad (25)$$

$$\text{for } \forall i' \in I, \forall j' \in J, \forall p' \in P_{i',j'}, \forall k' \in K$$

where  $c_{l_{p',n_{p'}}}(e_{l_{p',n_{p'}}}(k'))$  is the non-travel time cost (unit: dollar) of link  $l_{p',n_{p'}}$  at time  $e_{l_{p',n_{p'}}}(k')$  generated by vehicle usage, distance cost, etc.  $\kappa_{l_{p',n_{p'}}}$  is the value of travel time on link  $l_{p',n_{p'}}$ .

$z_{l_{p',n_{p'}}}(e_{l_{p',n_{p'}}}(k'))$  is the total number of containers on link  $l_{p',n_{p'}}$  at time  $e_{l_{p',n_{p'}}}(k')$ . All the data required in (25) can be obtained directly or computed approximately from the simulation model.

Then the marginal costs of a route is the sum of the time dependent link costs if a link cost is set as,

$$c_l(k') + \kappa_l w_l(k') + \kappa_l \frac{z_l(k')}{\Delta t} \frac{\partial w_l}{\partial y_l}(k') \quad (26)$$

where  $c_l(k')$  is the non-travel time cost (unit: dollar) of link  $l$  at time  $k'$  generated by vehicle usage, distance cost, etc. Therefore the problem of finding the routes with minimum marginal costs can be converted into an elementary time-dependent shortest path problem where the link costs are set as in (26). The shortest path algorithms in references (8-11) can be applied to find the new route with minimum marginal cost  $p_r$  in equation (19).

### 3.3 Load Balancing Algorithm

For each OD pair where a new route was found, the load balancing algorithm redistributes the freight loads among  $P_{i,j}(k)^{(m+1)}$  containing the current used routes and the new route based on the marginal costs for the OD pairs with the new found routes. One possible way to do the loading balancing is moving the freight loads between two routes iteratively until the marginal costs of all the routes are equal as done in [25]. However this load balancing algorithm faces a slow convergence problem for large demand sizes or large network sizes. Here we propose a load balancing algorithm with quicker convergence based on solving a linear programming problem using an auxiliary routing solution  $X_{Aux}^{(m)}$ ,

$$\min TC(X_{Aux}^{(m)}) \quad (27)$$

$$\text{where } \varphi_{i,j,k}^p(X_{Aux}^{(m)}) = \varphi_{i,j,k}^p(X^{(m)}) \text{ for } \forall i \in I, \forall j \in J, \forall p \in P_{i,j}, \forall k \in K \quad (28)$$

subject to

$$\sum_k \sum_{p \in P_{i,j}^{(m+1)}} x_{i,j}^p(k) = d_{i,j}, \text{ for } \forall i \in I, \forall j \in I \quad (29)$$

$$x_{i,j}^p(k) \geq 0, \text{ for } \forall i \in I, \forall j \in J, \forall p \in P_{i,j}, \forall k \in K \quad (30)$$

given  $d_{i,j}$  for  $\forall i \in I, \forall j \in J$

Then the new routing decision  $X^{(m+1)}$  can be generated using a step size method  $\alpha_m \in [0,1]$ ,

$$X^{(m+1)} = X^{(m)} + \alpha_m \left( X_{Aux}^{(m)} - X^{(m)} \right) \quad (31)$$

The most widely applied method of step size selection is the Method of Successive Averages (MSA) in which the step size is selected as  $1/(m + 1)$  as in the user equilibrium algorithms in [18-20]. Although MSA works well for small scale networks, its convergence is slow for large networks. In this report, we decide the optimal step size by solving the following optimization problem (32) in which the total cost of a new possible routing decision is evaluated by running the simulation model. Due to the fact it is time consuming to evaluate all the possible step sizes in the feasible set, we build a discrete set of candidate steps sizes to be evaluated.

$$\alpha_m = \arg \min_{\alpha_m \in [0,1]} TC \left( X^{(m)} + \alpha_m \left( X_{Aux}^{(m)} - X^{(m)} \right) \right) \quad (32)$$

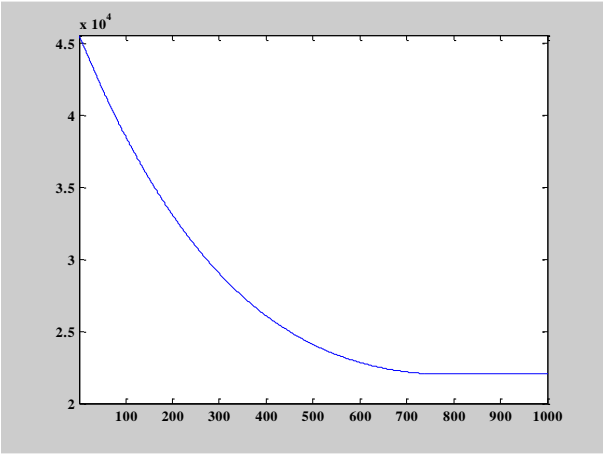
We next compare three different step size selection algorithms (i.e., enumeration approach in which one container is added to the minimum cost route during iterations, MSA, and the optimization approach using (32)) on a simple example. There are three possible routes connecting one OD pair whose characteristics and conditions during three time intervals (one-hour each in length) are shown in Table 1.

Table 1: Route Characteristics and Traffic Conditions of Simple Example

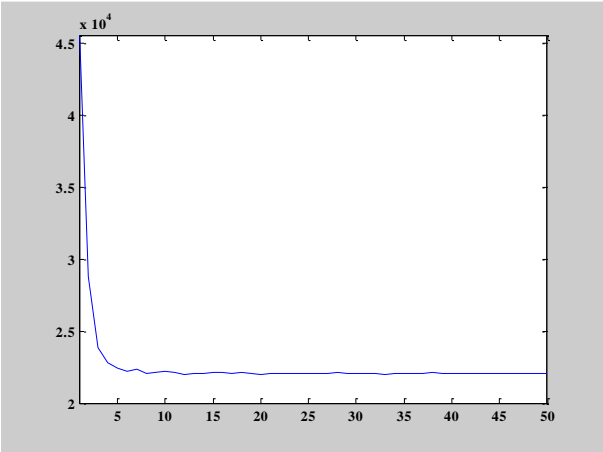
Time Interval	Route	Length (mile)	Capacity (veh/hour)	Current Demand (veh/hour)	Travel Time (min)
1	1	12	1100	1200	24
	2	10	1000	1000	18
	3	11	1050	1500	31
2	1	12	1100	1000	20
	2	10	1000	950	17
	3	11	1050	1100	21
3	1	12	1100	800	17
	2	10	1000	900	17
	3	11	1050	700	15

The number of vehicles between this OD pair is 1200 and the total cost is the sum of the travel

times in minutes of all the vehicles. Figure 2 shows the convergence of the three step size algorithms. The x-axis is the iteration number and the y-axis is the total cost for that iteration. The required numbers of iterations to stop for the three algorithms are about 750 for the enumeration approach, 20 for the MSA algorithm, and 6 for the optimal step size algorithm in (32). Therefore, the optimal step size method provides the better convergence speed although the three algorithms find the same optimal total cost.

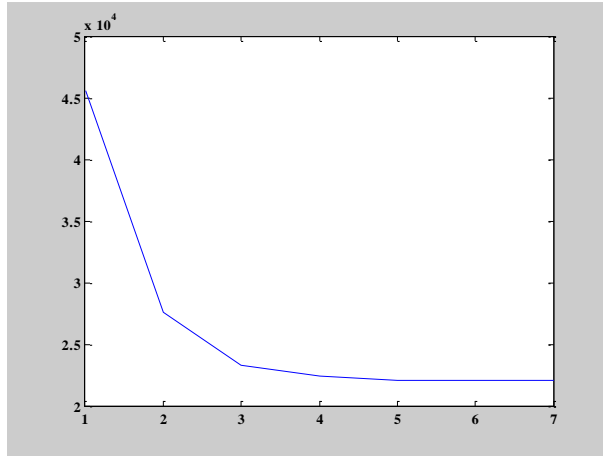


a) Enumeration Approach;



b) MSA Algorithm;





c) Optimal Step Size Algorithm

Figure 2: Performances of different load balancing algorithms

## 4 Evaluation Results

### 4.1 Simulation Models

The simulation models used in the COSMO approach of this report consist of a macroscopic road network model and a rail simulation model. We use the macroscopic traffic simulator VISUM to model the traffic flow in the road network shown in Figure 3 (an area that includes LA/LB ports and adjacent road network) as it is computationally faster than a microscopic model. The nodes in the road network model are road intersections or junctions connected by links that describe the street and freeway segments. The inputs including passenger and freight traffic for the road network are expressed as number of trips between zones that are origins and destinations within the road network. We assume that the trucks can only carry one container in the model so the number of truck trips between each OD pair will be the number of containers to be delivered. Historical passenger traffic data of year 2012 that are obtained from the Southern California Association of Governments (SCAG) are used to tune the simulation models. Since the data is only available for a portion of the links in the selected region, dynamic traffic assignment is used to estimate volumes for other network links.

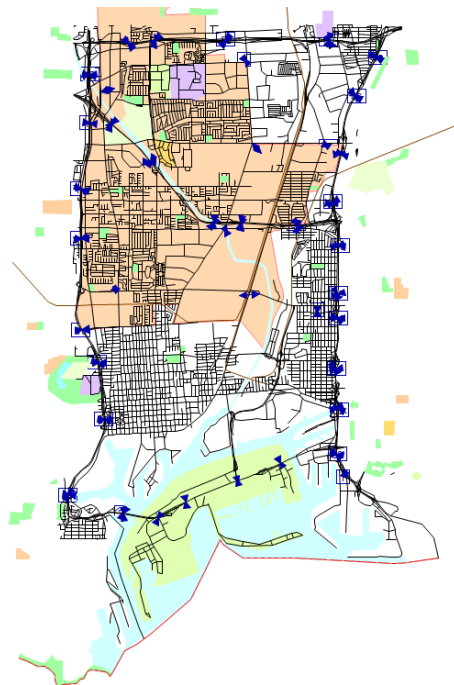


Figure 3: Road network simulation

For the rail simulator, we use the railway simulation system of Lu et al. in [29] which was developed based on the ARENA simulation software. The rail simulator is used to evaluate the dynamical train movements for complex rail networks. The track network is divided into different segments based on their speed limits, length, and locations. Then, an abstract track graph is constructed with these segments. The inputs for the rail simulator are the passenger and freight train schedules including their planned departure times, origin stations, and destinations. Then the train movements in the track network are simulated to calculate the travel times and delays of all involved trains.

The integration of the two models has been realized by sharing the OD demands and simulation outputs. The road network simulator sends the freight traffic with the information of the required departure time, and origin and destination stations that will be delivered through trains to the rail simulator. Then, the rail simulator creates the freight train schedule based on the train capacity and simulates the train movements with the planned passenger trains together to output the predicted train arrival times. After receiving the outputs of the rail simulator, the road network simulator will generate necessary truck flows to dispatch containers from the rail stations to their final destinations.

## 4.2 Case Study and Results

We evaluated the routing between six main destinations (D1 – D6) and three terminals (A, B, C) in the region with different demands as shown in Figure 4. We assume that there are five trains with homogenous capacities of 50 containers between the port terminals and two rail stations nearby the destinations. The average weight of all the containers is assumed to be 25 tons and transportation costs per unit (price/ton-mile) are estimated to be 8 cents for road and 3 cents for railway. The time value of a road link is set to be 40 dollars per truck hour and the time value of a rail link is 100 dollars per train hour. The demands of the six destinations are provided in Table 2. The total demands are equally supplied from three shipping companies (SC). Three shippers communicate their load demands to a coordinator who runs the COSMO approach to generate routes for their demands by minimizing the overall cost.

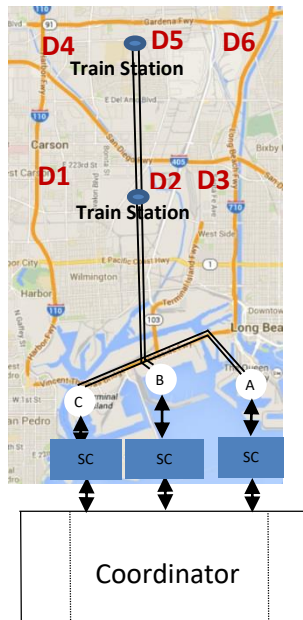


Figure 4: Region of study

Table 2: Demands of Destinations

Destination	1	2	3	4	5	6
Supply from A	0	60	400	0	0	560
Supply from B	0	390	0	0	630	0
Supply from C	350	0	0	600	70	0
Total Demand	350	450	400	600	700	560

Three traffic conditions are evaluated that are normal traffic, congested traffic and congested traffic with accident. In the normal traffic, the passenger traffic is set using the average daily volumes while in the congested condition the passenger traffic is increased by 50% above the average volumes. In the third case, lane closures are introduced on two locations on main freeways I-710 and I-110 causing the capacities of two freeway links to reduce by a half during the congested traffic condition.

Table 3: Simulation Results of Three Cases

Case	Normal Traffic	Congested Traffic	Accident Traffic
Average Cost of Initial Solution (Dollars/Container)	119.4	170.9	172.5
Average Cost of Final Solution (Dollars/Container)	66.3	79.5	80.5
Percentage Improvement	44.5%	53.4%	53.3%

Table 3 shows the average cost transferring all containers from their origins to the assigned destinations via the multimodal transportation network for the three traffic cases (normal traffic, congested traffic, and congested traffic with accident) for the case where the initially computed minimum cost routes are used and the case where the COSMO approach is used to do load balancing. As shown in the table, the COSMO approach leads to substantial improvement reduction especially in the case of unpredicted events such as congestion and accidents an initial solution did not anticipate.

## 5 Conclusion and Summary

We consider the multimodal freight routing problem by exploiting the effect of freight on traffic flow that can be predicted by more accurately using real time complex simulation models than the traditional use of simple mathematical models. We developed an approach we referred as to the COSMO approach that involves a feedback loop which includes a real time simulator, optimization and load balancing modules together with stopping criteria. The purpose of the load balancing module is to distribute the freight loads in the network in a way that leads to minimum cost routes despite the dependence of load on the route cost. The optimization module calculates the minimum cost routes based on the estimated network states generated by the real time simulation models. The iterations lead to reductions in the overall cost till certain stopping criteria are met in which case the final freight routing solution is ready to be applied to the actual system. The performance of the proposed approach is demonstrated with an example in the Los Angeles/Long Beach area. The computational results demonstrated the effectiveness of the proposed approach in reducing average costs under different traffic conditions that involve normal and unpredictable operations.

## 6 Implementation

A possible application of load balancing developed in this project is in the logistics planning or management of freight area. The proposed methodology can assist with making better freight routing decision that reduces the total delivery cost by considering the impact of freight traffic on transportation network.

The implementation of the proposed approach will require simulation software tools such as VISUM and ARENA as well as suitable programming tools such as C++, MATLAB, etc. It also requires access to the freight demand data from suppliers such as the number of containers to be transported together with origin to destination data and time frame.

## References

1. B. Jourquin, M. Beuthe. Transportation Policy Analysis with a Geographic Information System: The Virtual Network of Freight Transportation in Europe. In *Transportation Research Part C*, Vol. 4, No. 6, 1996, pp. 359-371.
2. F. Southworth, B. E. Peterson. Intermodal and International Freight Network Modeling. In *Transportation Research Part C* Vol. 8, 2000, pp. 147-166.
3. D. Rondinelli, M. Berry. Multimodal Transportation, Logistics, and the Environment: Managing Interactions in a Global Economy. In *European Management Journal*, Vol. 18, No. 4, 2000, pp. 398-410.
4. P. Nijkamp, A. Reggiani, W. F. Tsang. Comparative Modelling of Interregional Transport Flows: Applications to Multimodal European Freight Transport. In *European Journal of Operational Research*, Vol. 155, 2004, pp. 584-602.
5. P. Arnold, D. Peeters, I. Thomas. Modeling a Rail/Road Intermodal Transportation System. In *Transportation Research Part E*. Vol. 40, 2004, pp. 255-270.
6. C. Macharis, Y. M. Bontekoning. Opportunities for OR in Intermodal Freight Transport Research: A Review. In *European Journal of Operational Research*, Vol. 153, 2004, pp. 400-416.
7. P. Modesti, A. Sciomachen. A Utility Measure for Finding Multiobjective Shortest Paths in Urban Multimodal Transportation Networks. In *European Journal of Operational Research*, Vol. 111, 1998, pp. 495-508.
8. A. Lozano, G. Storchi. Shortest Viable Path Algorithm in Multimodal Networks. In *Transportation Research Part A*. Vol. 35, 2001, pp. 225-241.
9. K. F. S. Abdelghany. Stochastic Dynamic Traffic Assignment for Intermodal Transportation Networks with Consistent Information Supply Strategies. Dissertation of The University of Texas at Austin, 2001.
10. S. Demeyer, P. Audenaert, M. Pickavet, and P. Demeester. Dynamic and Stochastic Routing for Multimodal Transportation Systems. *IET Intelligent Transportation Systems*, Vol. 8, No. 2, 2014, pp. 112-123.
11. T. Pajor. Multi-Modal Route Planning. Thesis of [Universitat Karlsruhe](http://www.kit.edu). 2009.
12. Lu Liu. Data Model and Algorithms for Multimodal Route Planning with Transportation Networks. *Dissertation of Technische Universitat Munchen*, 2010.
13. D. Kirchler. Efficient Routing on Multi-modal Transportation Networks. *These of DOCTEUR DE L'ECOLE POLYTECHNIQUE*, 2013.
14. L. Castelli, R. Pesenti, W. Ukovich. Scheduling Multimodal Transportation Systems. In *European Journal of Operational Research*, Vol. 155, 2004, pp. 600-613.



15. H. Ham, T. J. Kim, D. Boyce. Implementation and Estimation of a Combined Model of Interregional, Multimodal Commodity Shipments and Transportation Network Flows. In *Transportation Research Part B*. Vol. 39, No. 1, 2005, pp. 65-79.
16. K. G. Zografos, K. N. Androutopoulos. Algorithms for Itinerary Planning in Multimodal Transportation Networks. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 9, No. 1, 2008, pp. 175-184.
17. J. Moccia, J. Cordeau, G. Laporte, S. Ropke, M. Valentini. Modeling and Solving a Multimodal Transportation Problem with Flexible Time and Scheduled Services. In *Networks*, Vol. 57, 2011, pp. 53-68.
18. J. E. Fernandez, J. D. Cea, A. Soto. A Multi-modal Supply-demand Equilibrium Model of Predicting Intercity Freight Flows. In *Transportation Research Part B*. Vol. 37, 2003, pp. 615-640.
19. X. Zhou, H. S. Mahmassani, K. Zhang. Dynamic Micro-assignment Modeling Approach for Integrated Multimodal Urban Corridor management. In *Transportation Research Part C*. Vol. 16, 2008, pp. 167-186.
20. K. Zhang. Dynamic Stochastic User Equilibrium: Models and Algorithms for Multidimensional Network Choices. *Dissertation of Northwestern University*, 2009.
21. Z.C. Li, W. H. Lam, S.C. Wong, and A. Sumalee. An Activity-based Approach for Scheduling Multimodal Transit Services. In *Transportation*, Vol. 37, 2010, pp. 751-774.
22. J. Dong, M. Makaiwi, N. Shafieirad, Y. Huang. Modeling Multimodal Freight Transportation Network Performance under Disruptions. Report of Iowa State University, 2015.
23. B. F. Russ, T. Yamada, J. T. Castro, and H. Yasukawa. Optimising the Design of Multimodal Freight Transport Network in Indonesia. In *Journal of the Eastern Asia Society for Transportation Studies*. Vol. 6, 2005, pp. 2894-2907.
24. T. Yamada, B. F. Russ, J. Castro, and E. Taniguchi. Designing Multimodal Freight Transport Networks: A Heuristic Approach and Applications. In *Transportation Science*, Vol.43, No. 2, 2009, pp. 129-143.
25. A. Abadi, P. A. Ioannou, M. M. Dessouky. Multimodal Dynamic Freight Load Balancing. *IEEE Transactions on Intelligent Transportation System*, Vol. 17, No. 2, 2016, pp.356-366.
26. N. Geroliminis; C.F. Daganzo. Existence of Urban-scale Macroscopic Fundamental Diagrams: Some experimental findings. In *Transportation Research Part B*, Vol. 42, No. 9, 2008, pp. 759-770.
27. *Traffic Assignment Manual*. U.S. Department of Commerce, Bureau of Public Roads, 1964.
28. G. A. Davis, H. Xiong. Access to Destinations: Travel Time Estimation on Arterials. *Technical Report MN/RC 2007-35*, University of Minnesota, 2007.
29. Q. Lu, M. Dessouky, and R.C. Leachman, Modeling Train Movements Through Complex Rail Networks, *ACM Transactions on Modeling and Computer Simulation*, Vol. 14, 2004, pp. 48-75.