Proceedings of the 2002 Winter Simulation Conference
E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds

USING SIMULATION MODELING TO ASSESS RAIL TRACK INFRASTRUCTURE IN DENSELY TRAFFICKED METROPOLITAN AREAS

Maged M. Dessouky
Quan Lu

Dept. of Industrial and Systems Engineering
University of Southern California
Los Angeles, CA 90089-0193, U.S.A.

Robert C. Leachman

Dept. of Industrial Engineering and
Operations Research
University of California, Berkeley
Berkeley, CA 94720, U.S.A.

ABSTRACT

We present a simulation modeling methodology to assess the rail track infrastructure in highly dense traffic areas. We used this model to determine the best trackage configuration to meet future demand in the Los Angeles-Inland Empire Trade Corridor Region. There are three major challenges in modeling a rail network in a densely trafficked metropolitan area. They are: (1) complex trackage configurations, (2) various speed limits, and (3) non-fixed dispatching timetables and routes between the origin and destination. Our proposed model has the ability to handle the above complexities in order to determine the best use of the rail capacity. Furthermore, our methodology is general enough so that it can be applied to other large scale rail networks.

1 INTRODUCTION

As global trade continues to increase, cargo traffic at the nation's ports continues to increase at dramatic levels. The annual rate of container growth in the Uinited States is around 6% (Vickerman, 1998). For example, the Ports of Los Angeles and Long Beach (San Pedro Bay Ports) are among the busiest ports in America. Booming trade with Pacific Rim nations has seen the annual trade in the two ports exceed 100 million tons. Railways form the major means to transcontinentally move these goods. Forecasts show that volume growth of high-value containerized freight is set to triple in the next 25 years. This rapid growth has already introduced congestion and threatened the accessibility and capacity of the rail network system in the Los Angeles Basin and other locations.

To partially address this rail traffic growth, innovative projects like the Alameda Corridor (Los Angeles County, CA) are being developed. The Alameda Corridor is a high-speed multiple-track line from the Ports to Downtown Los Angeles (Leachman, 1991). However, the Alameda Corridor does not address increased rail traffic east of Downtown Los Angeles (see Figure 1) Therefore, there is a need to develop tools to model these types of complicated rail networks.

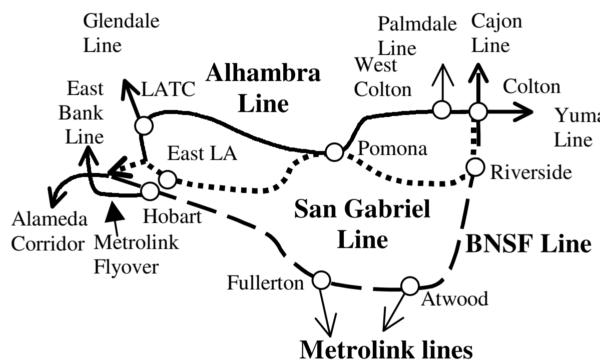

Figure 1: Railway Networks of the Los Angeles-Inland Empire Trade Corridor

The rail network in this area is extremely complex because of the need to make maximal use of the capacity of the rail network. For example the trackage configuration consists of single, double and even triple track. Furthermore in some high traffic zones or line connection zones the rail network cannot be considered as a combination of portions of single, double or triple track. Figure 2 gives an example of a complex junction. It is an alternative track configuration for the West Colton Junction in the Year 2020. As the diagram shows, there is a multitude of ways to cross over from the Alhambra line to the Palmdale line. In some of the paths, the main line may be blocked by a crossover train. As opposed to the trackage configuration in rural areas, these types of trackage configurations are quite common in metropolitan areas.

Another complicating factor in urban rail networks is the existence of multiple speed limits at different points in the network. For example, in Figure 2 the detour tracks require imposing lower speed limits than the straight line tracks. Besides physical contours, other reasons for changing speed limits may be due to crossovers and safety

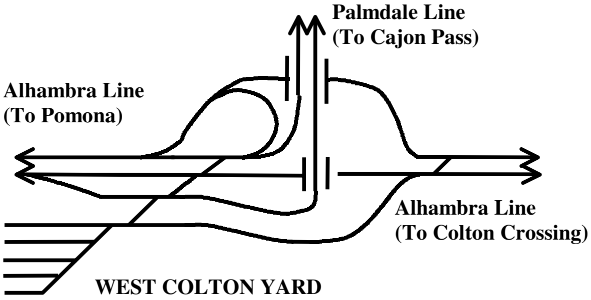

Figure 2: West Colton Flying Junction

considerations. In the case when multiple speed limits are considered, a central issue is to determine the fastest the train can travel at each instance of time without violating the speed limits considering the train's acceleration and deceleration rates.

Some other complicating factors include the existence of dedicated tracks and train priorities. To eliminate unnecessary conflicts, some tracks are dedicated only for trains running in one direction in the rail network just like a one-way road for vehicle traffic. Therefore, the dedicated and non-dedicated tracks must be distinguished in the network. To determine whether or not to bypass trains and the dispatching sequence when there are multiple trains ready in a station, priority is set to each type of freight and passenger trains. Generally we give passenger trains higher priority due to them operating on a fixed schedule.

Because trains with higher priority can bypass trains with lower priority, trains with the same origin and destination may run through different paths. A path indicates the specific segments of the track that the train will utilize. Since many portions of the network include double and triple-track segments, the alternative paths may be adjacent to one another. They may differ only on the side of the track that the train will travel and the location of any possible crossover points. This flexibility of routing between origin and destination stations is very susceptible to deadlocks. Hence, developing a central dispatching algorithm that decides the movement of each train in the network without causing any deadlock is necessary and important. In fact, for a general trackage network, the problem of determining the optimal dispatch times that minimize train delays and ensures deadlock-free operations is NP-hard.

Our goal is to develop a simulation methodology for modeling train movement in the rail network in the Los Angles-Inland Empire Trade Corridor in context of all the above mentioned complicating factors: multiple-track configurations, priorities, multiple speed limits, flexible routing and track dedication. The simulation methodology is used to analyze and evaluate the operational feasibility of various alternatives proposed to improve the rail network capacity in this area to effectively hand the increased rail traffic by the Years 2010 and 2020.

There has been some prior work in simulation modeling of rail networks. Dessouky and Leachman (1995) developed a simulation modeling methodology for either strictly single-track or double-track rail networks consisting of a single speed limit without considering deceleration rates. Petersen and Taylor (1982) present a structured model for rail line simulation. Higgins and Kozan (1998) proposed an analytical model to quantify the positive delay for individual passenger trains and track links in an urban rail network.

The modeling methodology presented in this paper differs from the previous work by considering multiple trackage configurations in the same rail network with multiple speed limits while taking into account the train's acceleration and deceleration rates. Even though the proposed simulation modeling methodology was applied to the rail network from Downtown Los Angeles to the Eastern Inland area, it can be applied to various situations to simulate rail networks with any kinds of topology, crossovers and speed limits.

2 SIMULATION MODEL

The simulation model is developed using the AweSim Simulation Language (Pritsker and O'Reilly, 1999), but may be implemented using any general-purpose simulation language. Train movement is a continuous process while the scheduling and dispatching of trains are triggered by discrete events. Therefore, our approach is based on a discrete event methodology. We approximate the continuous motion of train movement by dividing the movement in small discrete steps.

The physical resources that we model are: rail junctions and track segments. Figure 3 illustrates the overview of the model structure.

The data are input using files from three categories: train schedule, train type and track network. The departing train entities at each station are generated based on an input train schedule data file. These entities are stored in the event calendar according to their scheduled departure time from the origin station. The entity with the earliest scheduled time will leave the event calendar at its scheduled time.

If this entity is a train entity, the central dispatching algorithm is called to decide whether this train should continue moving or begin to decelerate to stop. A train begins to decelerate to stop either when some necessary trackage or junction resource is not available or a continue movement of the train may cause deadlock. If the train is stopped, the train entity is placed in a queue to wait for an available resource or the possible deadlock situation to be resolved.

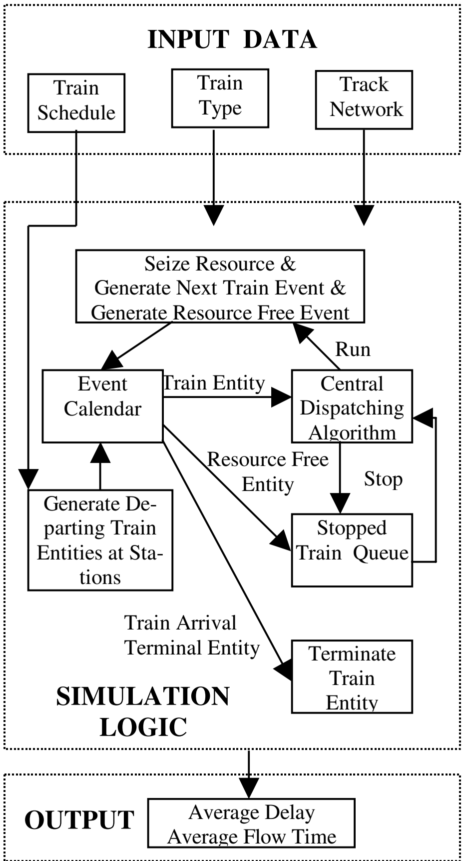

Figure 3: Overview of Model Structure

If the train is informed to move, the central dispatching algorithm determines the following:

- The successor node of the next movement
- The length the train travels within the node
- The speed of travel and possible change-of-speed points

Then, the train entity will seize all the necessary resources, schedule the resource free event to release the resources it no longer needs during this movement, and an event is scheduled to represent the time the train finishes movement.

If the entity is a resource free entity, all the train entities in the stopped train queue are checked to see whether this released resource can trigger a train movement for one of the stopped trains. The triggered train entity in the queue is the one with the highest priority and longest waiting time. Its movement will be determined by the central dispatching algorithm in the same manner as previously described.

Finally, if the entity out of the event calendar is an arrival terminal entity, the statistical information of this train entity will be recorded and the entity will be terminated from the system. When the simulation finishes, the primary outputs are the average delay and flow time of the trains.

2.1 Input Data

Each train type is associated with the following characteristics:

- Train length
- Maximum velocity
- Acceleration rate
- Deceleration rate

Each train schedule contains the following information:

- Departure Node
- Destination Node
- Train type (e.g. intermodal, bulk, carload, etc.)
- Priority number
- Time range list
- Average number of departure trains
- Interval distribution type

Each schedule is associated with a time range list. Each item in the list represents a specific time interval. The average number of departing trains during this interval is also an input to the model. The interval distribution type determines how to calculate the exact departure time of each train. In our model, for all the passenger trains, the time between arrivals of each train in one time range are based on a fixed schedule; for the freight trains, the time between train arrivals in one time range are exponential random variables.

2.2 Network Construction

The physical rail network consists of rail junctions and track segments. A rail junction is typically used for train cross-over movement in a rail network. One idea behind the modeling approach is to divide the physical track into segments as in Dessouky and Leachman (1995). A segment is the minimum unit shown in the proposed simulation model and each segment is represented as a unique resource with capacity one. A track segment has the following two characteristics:

- Travel in each segment is restricted by one speed limit.
- The length of the segment is no longer than the maximum train length.

The first characteristic is not restrictive since there is no limit on the minimum length of the segment. Hence, the definition of the segment is sufficiently generic to model any physical trackage configuration. However, having many small track segments will increase the number of resources in the simulation model and the computational run time of the model. On the other hand, since we restrict the capacity of each segment to be one, too large of a segment definition will increase the headway between trains, needlessly decreasing the capacity of the network. Thus, the second characteristic restricts the maximum size of the segment to be the maximum train length.

As we mentioned before, due to the complexity of the trackage configuration in an urban rail network, the rail network cannot be considered as a composition of different portions, where each portion is only a single or double track system. Instead, we have to think about the rail system as an entire general network, which comprises of nodes and arcs.

Each node in our network defines a combination of one or more contiguous segments. And each node has two ports: port 0 and port 1. Port 0 indicates the starting point of traveling on the contiguous segments of the node from one direction. Port 1 indicated the starting point of traveling in the opposite direction of Port 0. Two distances locate each segment in a node:

- The length of the segment itself, and
- The distance from the end of segment to Port 1 of this node.

Note that the length of the node equals the sum of all the segments' length of this node.

The nodes are connected by arcs, which represent movement from one node to another. Arcs may include junctions or not. All the arcs in our network are undirected and have zero length. Therefore, the total travel distance of a train in the network equals to the sum of the length of the nodes it visits. We illustrate the network concept in Figure 4 for the West Colton Flying Junction given in Figure 2.

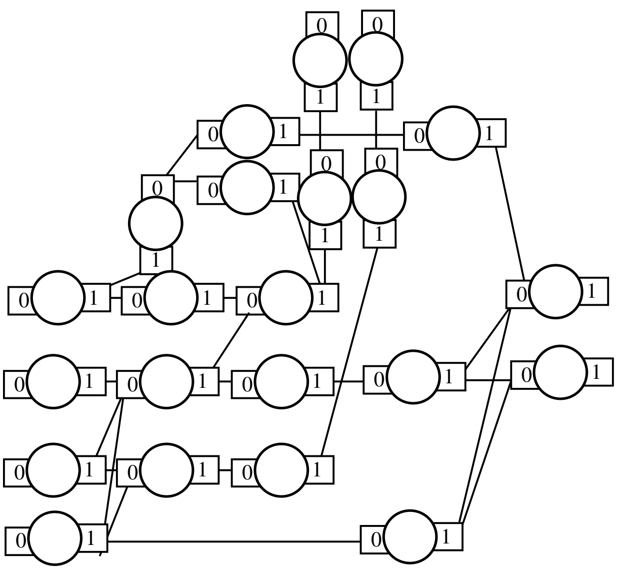

Figure 4: Simulation Network for the West Colton Flying Junction at Figure 2

The network is defined and stored in a file. The following is part of the network definition file. Each node contains one or more segments or junctions. Each arc connects nodes and may contain one or more junctions. The '*' following a junction name denotes that this junction is needed for a train passing it, but no junction speed limit needs to be applied.

```
;Node   Resource    Distance To    Resource
;ID     Name        Port 1         Length
;-----  --------    ------------   -----
   0    Segment10       0          1.52

   1    Segment9        1.65       0.01
        Segment3        0          1.65
        Juct1_1*        1.35       0
        Juct_1*         1.15       0
        Juct_2*         0.1        0

; Arc Definition
; From  From  To   To   # of      Names of
; Node  Port  Node Port Resources Resources
; ----- ----  ---- ---- --------- -----------
   1     1     3    0      1      Juct4_1*
   1     1     4    0      2      Juct4_1
                                  Juct4_2
```

2.3 Central Dispatching Algorithm

The central dispatching algorithm has two main tasks,

- Determine the optimal run times for a train under multiple speed limits
- Determine the next train movement.

In our simulation methodology, each node may contain multiple track segments and junctions. Since each track segment or junction has its unique speed limit, one node can include multiple speed limits. The question is how to accelerate and decelerate a train properly to let it cross a distance with multiple speed limits in a shortest amount of time. When the acceleration and deceleration rates are infinite, the optimal solution is always move the train at the velocity of the speed limit. When the acceleration and deceleration are finite, the optimal speed at each instance of time is a complicated combination of the acceleration and deceleration rates and the uniform motion of the train. Figure 5 shows a sample optimal speed function, given as the dotted lines, under multiple speed limits. The optimal speed is the dotted line that gives maximum area under the curve considering the acceleration/deceleration rates.

Based on the foregoing, the following two conditions below must be met in order for a train to move to a successor node:

- All track segments' resources of the successor node must be currently available as well as any possible connecting junction resources.
- A movement of a train to this successor node must not create a deadlock.

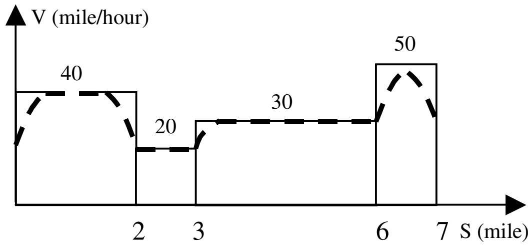

Figure 5: A Sample Optimal Development of a Train under Multiple Speed Limits

A successor node is considered *available* if the two above conditions are met. If more than one successor node is available, some heuristic criteria are used to select one of them as the successor node. The heuristic must consider the following three factors.

- The maximum priority difference between the current train and the immediate successor train running in the same direction if one exists.
- The maximal number of trains running in the same direction along the path from the successor node to the train's destination node.
- The minimum travel time for the current train from the successor node to its destination node assuming there is no downstream conflicting traffic ahead of the current train.

Each above factor emphasizes a different aspect of operating efficiency. If priority is based on the speed of the train, the first condition allows higher speed trains to bypass slower speed trains. The second condition maximizes the number of trains moving in the same direction for a given path, thus freeing other paths for opposing moving trains. The third condition minimizes the travel time for the train.

Note that we cannot wait until the head of the train reaches the end of the node to make a decision on whether a train should stop or not since we need to account for decelerating time, and some distance is necessary for a running train to fully stop. Therefore, the dispatching algorithm must be applied before the head of the train reaches the end of the node. In our modeling framework, we introduce the concept of a "stop checking point" which is the point of applying the dispatching algorithm to decide whether a train should stop or not. We use multiple stop checking points along a track segment instead of only one stop checking point. This can prevent unnecessary stops for the trains in the model.

When a train entity is at a stop checking point and there exists more than one available successor node, the proposed dispatching algorithm will choose the node indicating the next train movement. This dynamic selection is dependent on the status of other trains in the network at this instant of time. Therefore, even trains having the same schedule may on different days follow different paths from their origin to destination.

Although this flexible routing improves the usage of the network's capacity, it is highly susceptible to deadlocks. A typical deadlock situation arises when two trains running in the opposite direction are routed to nodes representing the same single-track simultaneously. Typical methods to resolve deadlock are:

- Restore the previous status until no deadlock will happen,
- Allow the preemption of the resource, and
- Design some routing algorithm that avoids the deadlock.

The first two methods don't work in our situation, because it is difficult and sometimes impossible for trains to move backward. Therefore, designing a deadlock-free and efficient dispatching routing algorithm is a core aspect in the central dispatching algorithm.

Our dispatching algorithm guarantees the avoidance of the deadlock. We illustrate our algorithm using the example shown in Figure 6.

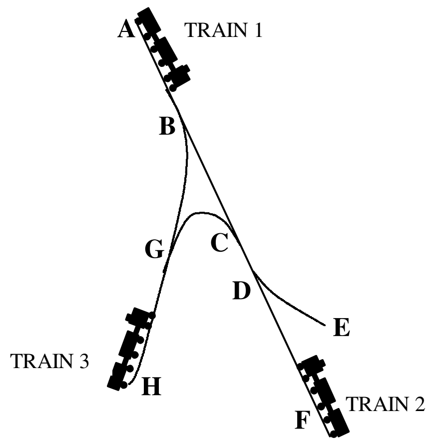

Figure 6: Example Illustration of Dispatching Algorithm

In Figure 6, there are three trains, train 1, train 2 and train 3. They are currently at locations A, H and F, respectively. The destinations of train 1, 2 and 3 are locations H, A and E, respectively. Assume that the length of segment BG is less than the length of Train 1, the length of segment BC is less than the length of Train 2 and the length of GC is less than the length of Train 3.

According to our dispatching algorithm, these trains will be dispatched using the following sequence. First, all three trains will continue moving. Then Train 1 and Train 2 will stop before location B and D respectively to wait until Train 3 crosses location G. When Train 3's tail passes location G, Train 1 will start to move to location H. Train 2 will start to move after the tail of Train 1 passes location B.

This dispatching sequence satisfies both the efficiency and deadlock avoidance.

3 SIMULATION RESULTS

The simulation model is used to assess the feasibilities of various alternatives proposed to improve the rail network's capacity in the Los Angeles-Inland Empire Trade Corridor.

Currently, there is 195 miles of track in this area. We divided this track into 330 segment resources and 178 junction resources. From these resources, we developed a network architecture consisting of 412 nodes and 593 arcs. Today, there are around 141.5 freight trains per day and 101 passenger trains per day that use this portion of the rail network. By 2010, these numbers are expected to increase to 278.2 freight trains per day and 227 passenger trains per day.

There are generally three strategies used in the alternatives to increase the capacities of the rail network.

- Expand the tracks, e.g. expand the current single track part to double track, and double track part to triple track etc.
- Grade separation at major crossings.
- Change the freight trains' routes.

Bottlenecks can be determined using our simulation modeling methodology, thus, identifying locations where additional trackage is needed to meet the increased demand. Table 1 shows some summary result output of the analysis. This set of results are based on a simulation run time of 100 days with clearing the statistics after 10 days. This took around 25 CPU minutes on a Pentium III 1 GHZ processor.

The first set of results show the train delays at the current volumes and trackage. The second set of results show train delays when volume reaches the expected 2010 levels using the current trackage configuration. Note that at the increased train volumes, the current trackage cannot sufficiently handle the higher traffic. Therefore, recommendations to increase the trackage at specific locations were made. For example, the following is one alternative solution for the year 2010.

- Expand the BNSF line from Commerce to Cajon to fully triple track.
- Expand the UP San Gabriel line from East Yard to Pomona to fully triple track.
- Expand the UP Alhambra line from Pomona to West Colton to fully double track.
- Add the Passenger train flyover at Pomona

The last set of results show the delays when the trackage is modified to handle the increased traffic.

Table 1: Simulation Results

	Current Trains in Current Trackage		2010 Trains in Current Trackage		2010 Trains in Improved 2010 Trackage	
	Ave Delay Time (min)	# of Trains per Day	Ave Delay Time (min)	# of Trains per Day	Ave Delay Time (min)	# of Trains per Day
Freight Trains	30.99	141.5	2430.8	195.1	26.23	195.1
Passenger Trains	12.23	101	106.2	163	9.24	163

4 CONCLUSION

Our train simulation modeling approach has shown to be an effective tool for analyzing rail capacity in the Los Angeles-Inland Empire Trade Corridor. The simulation analysis played an essential role in evaluating the alternatives and improving on them by identifying bottlenecks. Furthermore, our methodology is general enough so that it can be applied to other large scale rail networks.

REFERENCES

Dessouky, M. and R. C. Leachman. 1995. A Simulation Modeling Methodology for Analyzing Large Complex Rail Networks, *Simulation* 65(2); 131-142.

Higgins, A. and E. Kozan. 1998. Modeling Train Delays in Urban Networks. *Transportation Science* 32(4); 346-357.

Leachman, R. C. 1991. Railroad Capacity and Operations Analysis for the Alameda Consolidated Transportation Corridor Project *Technical Report*, Berkeley, CA.

Petersen, E. R., and A. J. Taylor. 1982. A Structured Model for Rail Line Simulation and Optimization. *Transportation Science* 16(2); 192-206.

Pritsker, A. B. and J. J. O'Reilly. 1999. *Simulation with Visual SLAM and Awesim.* John Wiley & Sons, Inc., New York.

Vickerman, M. J. 1998. Next-Generation Container Vessels: Impact on Transportation Infrastructure an Operations. *TR News* 196; 3-15

AUTHOR BIOGRAPHIES

MAGED M. DESSOUKY is an Associate Professor in the Department of Industrial and Systems Engineering at the University of Southern California. He received his Ph.D. in Industrial Engineering from the University of California, Berkeley, and M.S. and B.S. degrees from Purdue University.

QUAN LU is a Ph.D. candidate in the Department of Industrial and Systems Engineering at the University of Southern California. His research interests are in the area of simulation and optimization.

ROBERT C. LEACHMAN is a Professor in the Department of Industrial Engineering and Operations Research at the University of California, Berkeley. He received the Ph.D. and M.S. degrees in Operations Research and A.B. degree in Physics and Mathematics from U.C. Berkeley.