Pergamon

Transpn Res.-A, Vol. 32, No. 4, pp. 279–295, 1998

Printed in Great Britain
0965-8564/98 $19.00 + 0.00

PII: S0965-8564(97)00009-8

PREDICTING ON-TIME PERFORMANCE IN SCHEDULED RAILROAD OPERATIONS: METHODOLOGY AND APPLICATION TO TRAIN SCHEDULING

SUSAN F. HALLOWELL
Operations Research Group, US Air, 2345 Crystal Drive, Arlington, VA 22227, U.S.A

and

PATRICK T. HARKER*
Department of Operations and Information Management, University of Pennsylvania, Philadelphia, PA 19104-6366, U.S.A.

(Received 29 December 1995; in revised form 20 February 1997)

Abstract—In previous work, the authors developed an analytical line delay model for analyzing rail line-haul operations, and validated the model as a predictive tool. This paper examines the application of the model as a prescriptive tool for the generation of train schedules. A unique feature of the model is that it incorporates dynamic meet/pass priorities in order to approximate an optimal meet/pass planning process. Extensive Monte Carlo simulations are conducted to examine the application of the model for adjusting real-world schedules to improve on-time performance and reduce delay. This empirical work represents the first attempt to investigate the impacts of the scheduling methodology on on-time line-haul performance. The problems are based on historical data from a major North American railroad.

Keywords: railroads, stochastic scheduling, computer assisted dispatching

1. INTRODUCTION

This paper considers the problem of scheduling trains on a partially double track rail line. On such a rail line, meet/pass interference is the major cause of delay (Smith *et al.*, 1990). Delays occur on a single track when a train switches to the side track to allow another train traveling in the same direction to pass. Delays also occur when a train switches to a side track to allow another train to pass. Scheduling is complicated by the uncertainty in line-haul operations. There is uncertainty regarding when trains depart and which train will be delayed when a meet or pass arises. This uncertainty makes it difficult to determine reliable schedules.

This research is motivated by the North American railroads' concern for freight service reliability and for capacity on a number of key lines. According to one marketing survey, a 1% improvement in the reliability of cargo delivery time could yield as much as a 5% revenue increase in several markets (Hertenstein and Kaplan, 1991). In North American railroad operations, there is certainly a link between line delay and service reliability for both Amtrak and Intermodal traffic, although this link is much less clear for general merchandise traffic because the unreliability of yard service times may overwhelm the unreliability of line travel times (Martland and Smith, 1990; Smith, 1990a; Martland *et al.*, 1993).

There are at least two operating strategies to improve service reliability and increase line capacity: master scheduling and real-time scheduling (Harker and Ward, 1991; Harker, 1995). With master scheduling, railroads would move towards a European approach; i.e. develop a detailed timetable for scheduled trains and slots for unscheduled trains, and then operate with strict adherence to these schedules. With real-time scheduling, railroads would use schedules more as

*Author for correspondence.

guidelines in making decisions as to how trains should operate. All major North American railroads are contemplating both these strategies, although the trend is to move towards a master scheduling strategy.

This paper addresses these new scheduled-based approaches to operating a railroad where there are needs for new tools and processes to determine reliable train schedules. Figure 1 presents a conceptual hierarchy of rail planning and control models developed by Jovanović and Harker (1991). Within the hierarchy, information flow passed downward specifies targets (goals) to determine control decisions. Information flow passed upward provides feedback on progress towards achieving goals. Given the physical resources and services defined at the strategic level, tactical scheduling of trains involves the effective allocation of resources, which entails the creation of a set of reliable schedules. Models at the on-line control and planning tiers in the hierarchy are viewed as supporting operations with the aim of making planning decisions that will achieve the targets for dispatchers and yard personnel. These targets need to be logically consistent in the sense that planning decisions exist that can achieve the targets with high probability. System-wide real-time scheduling is further decomposed into dispatching, train control and yard operations, where decisions are made regarding the movements of cars, locomotives, switch engines, crews, etc.

We consider the application of the analytical line delay model (LDM) and method for calculating likely train arrival and departure times (the *target time generator* or TTG) developed by Hallowell and Harker (1996). The LDM/TTG is meant to be used to adjust the targets (schedules) for a given rail line. Adjustments arise in both the tactical and real-time scheduling process, whenever conditions change on the line. For instance, a train may be re-routed or an additional train may be added. Owing to the interdependency of resources, disruptions to one train's targeted times may propagate throughout the rail network, requiring the adjusting of other trains' targeted times. While this work focused on adjusting the targets for a rail line, the long term goal is to incorporate the LDM/TTG within a higher-level network scheduling tool; e.g. see Kraay and Harker (1995) for a real-time network train scheduling model.

1.1. Overview

The first goal of this paper is to examine the performance of the LDM/TTG, i.e. are there meet/pass decisions which can achieve the targets prescribed by the LDM/TTG with high reliability? We

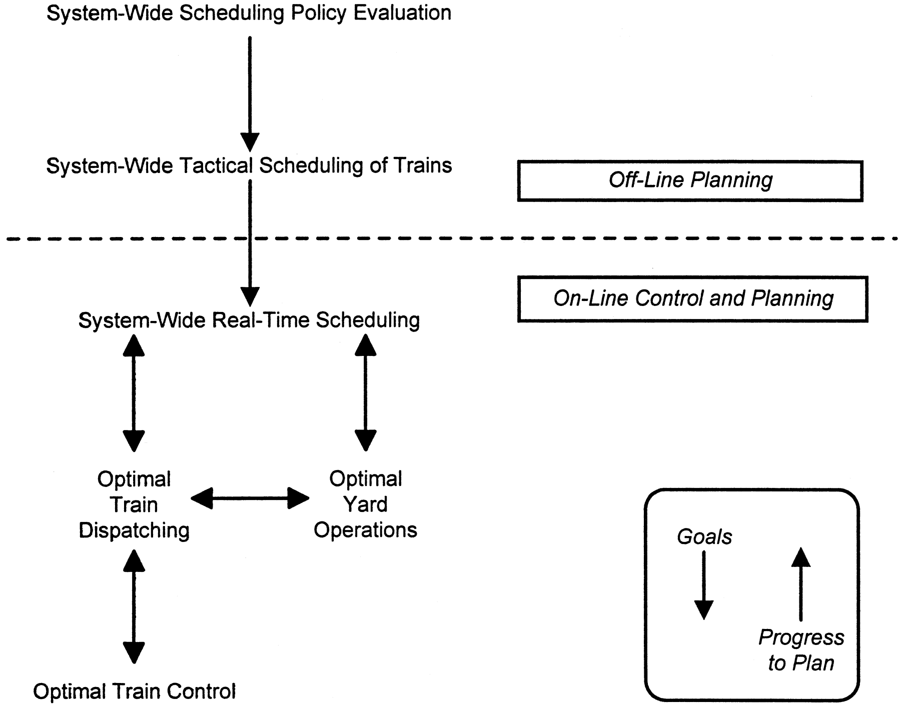

Fig. 1. Hierarchy of rail operations models.

also consider the possible computational advantage of the LDM/TTG analytical approach over a traditional simulation-based approach. If every line in a rail network required a detailed simulation, the network problem may become intractable. The analytical approach provides relatively fast adjustments to the targets which are expected to reduce the need for detailed simulations.

The second goal is to examine the potential impact of using the LDM/TTG to tighten the targets. This analysis is a first attempt to investigate the impacts of the scheduling methodology on line-haul performance, in particular, on-time performance. In an earlier study using existing schedules from a major railroad (Hallowell, 1993), we found that trains could arrive consistently early under an optimal on-time performance-based meet/pass planning policy, i.e. there was excessive schedule slack (padding), particularly for bulk and mixed freight trains' schedules. In this paper, comparative analysis with traditional schedule updating policies examines the impacts of adopting a scheduling methodology aimed at 'tightening' the targeted times. Our results suggest that there are opportunities to achieve a greater precision of on-time performance by reducing slack with the LDM/TTG. In the long term, reductions in slack may result in increased productivity and this may translate into decreased operating and fixed costs.

Extensive Monte Carlo simulations are conducted to test the LDM/TTG. Train movements are simulated over a period of several days for a given set of targets. The input uncertainty is in the departure times; the time a train is ready to depart from its origin point within the line is a major source of uncertainty in the dispatch environment (Lang and Martland, 1972; Hallowell *et al.*, 1991). A simulation model is used to determine target performance as measured by the sample mean and variance of arrival time and delay. In order to examine the impact of the scheduling methodology, we compare the performance of the TTG-adjusted targets to the performance of 'Base Case' (real-world) targets. Sensitivity analysis is performed by varying traffic volumes and departure uncertainty.

The tests consider the case of an on-time-performance-based meet/pass planning policy. Optimal on-time performance-based meet/pass planning systems are expected to yield substantial improvements in on-time performance and utilization of track, locomotives, and rolling stock (Smith *et al.*, 1990; Smith, 1990b). It should be noted, however, that the state-of-the-art (1995) computer-assisted dispatching system developed by Union Switch and Signal and installed by Union Pacific and CSX railroads is fixed priority-based.

Within the simulation there is an optimal meet/pass algorithm to determine meet/pass decisions. The algorithm is part of the patented scheduling tool called SCAN II (Schedule Analysis System) developed at the University of Pennsylvania and used by the Burlington Northern Railroad (Jovanović and Harker, 1990, 1991). The algorithm's objective is to determine meets and passes which minimize total weighted tardiness (cost associated with non-negative lateness at scheduled points). This objective is consistent with scheduled rail operations.

A unique feature of the LDM/TTG is that it incorporates dynamic meet/pass priorities in order to approximate on-time performance-based meet/pass decisions. Actual meet and pass conflicts are not resolved by the LDM/TTG. Rather, the model predicts the expected delay. The resulting expected delays are then used to update the targets. The LDM/TTG could also be used to adjust the targets under a fixed dispatch priority case, although this is not considered here.

The remainder of this paper consists of four sections. The next section describes how the LDM/TTG is used to adjust the targets. In Section 3, we present the methodology of the empirical tests, and in Section 4, we present the test results. Section 5 summarizes the main results of the research and presents some ideas for future work.

2. THE LDM/TTG METHOD OF ADJUSTING THE TARGETS

In the train schedule adjustment process, we are given a set of initial targets, which we refer to as the targeted times of departure (TTD) and arrival (TTA). We assume all trains have targets at major schedule points where they enter and exit the line (the ends of dispatching territories, yards, junctions, etc.). Trains may also have targets at intermediate points within a line, e.g. a train may have a scheduled dwell time to pick up or drop off cars, passengers, or crews. The goal is to adjust the targets in a way that will achieve a given reliability objective.

In practice, we expect railroads would want to determine targets that achieve a given percentage (e.g. 80, 85, 90%) of on-time arrivals within a given arrival time window. For instance, one long-term operating goal for service at CSXT is to determine line-haul targets that achieve 98% arrivals around a 1 or 2 h window for general merchandise trains.

Currently, adjusting the targets is a manually intensive, heuristic process. Schedulers may use primitive tools (e.g., pens and ink boards) or sometimes graphical line simulation tools to determine when and where trains will encounter meet/pass interference. In the manual process, the scheduler iterates sequentially through pairs of trains which are expected to meet or pass. When two trains interfere, slack is added to the trains' target arrival times to cover the anticipated delay.

The scheduler attempts to add enough slack to ensure a given reliability, but not any excess. In some cases adding extra slack will improve the reliability of shipment delivery time. For instance, cars have a higher probability of making connections if they arrive early at a yard (Martland, 1982; Martland and Smith, 1990). On the other hand, too much extra slack may cause down-line congestion and inefficient use of resources. A chief problem with early arrivals is that they often cannot be served immediately, but rather must wait for an available server, i.e. terminal/yard facilities, a fresh crew, etc. For example, a train which arrives early to a classification yard may have to wait on the main line outside the yard, causing delays to other trains ready to depart. Another problem with excessive slack is that it is inefficient. Generally speaking, slack is directly related to cycle times for cars and locomotives, infrastructure costs, and crew costs. Hence, one of the objectives of scheduling is to limit slack.

The LDM/TTG is a tool aimed at enabling precision scheduling. It is an analytical line delay model in which we first estimate expected delay and then adjust the targets based on the expected delay. The overall structure of the LDM/TTG is presented in Fig. 2. Details of the model are given in Hallowell and Harker (1996). The estimate of delay depends on the dispatch policy, since it is this which determines which train will be delayed in a meet/pass conflict. In our tests, the dispatch policy is minimize total weighted tardiness costs, which is specified by the weights per train. Delay estimates are also determined by the input targets under an on-time performance-based dispatch policy. Thus, the inclusion of the targets as an explicit input is significant.

The output of the LDM/TTG are the estimates of mean delay and the set of adjusted targets. The delay estimates are derived using a set of equations of fixed point form developed to accurately model line-haul operations. The resulting adjusted targets are derived by adding the expected departure time to the estimated mean delay.

Under an on-time performance-based dispatch policy, repeated applications of the model should be taken until convergence for the targets is achieved, to ensure that the resulting targets

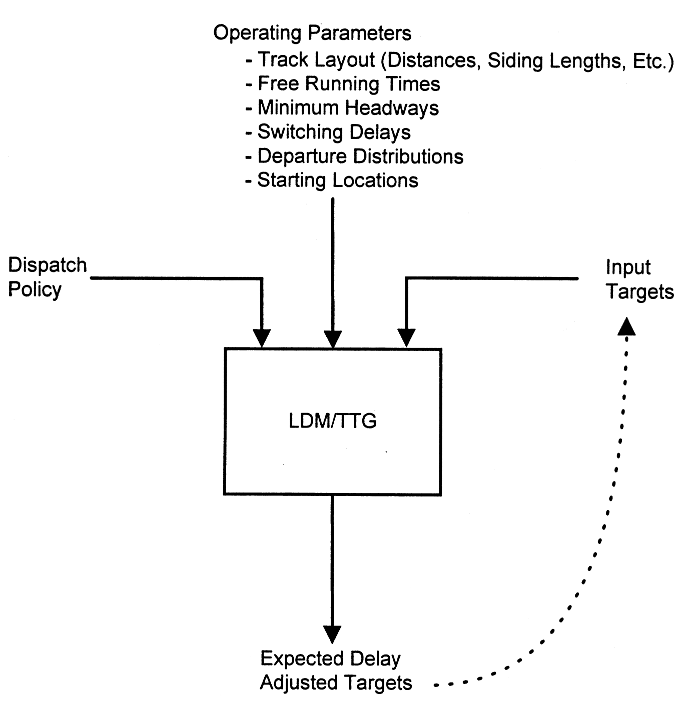

Fig. 2. Structure of the LDM/TTG.

are consistent. That is, the output targets may be used as inputs to another iteration of the model, and this procedure may be repeated until convergence is achieved (the targets do not change by more than some tolerance). We have chosen to test a heuristic that entails one iteration of the LDM/TTG procedure. In preliminary tests, we found that this provides fairly accurate solutions within a reasonable amount of time. Further work is required to consider the potential benefits that may be achieved by an iterative method.

The LDM/TTG is designed to prescribe targets so that trains arrive early 50% of the time. This is because the adjusted targets are based on estimated mean delay. Admittedly, railroads may want some trains to have a higher probability of being early rather than late. For these trains the LDM/TTG could be used to derive a lower bound for the target arrival time.

Several benefits are expected to be derived from the LDM/TTG schedule adjustment tool. We expect it would improve the productivity of schedulers by automating a fairly tedious part of their job. Improved productivity would allow them more time to focus on refinements to the schedule, and therefore increase the quality of their solutions. A schedule adjustment tool would also be of benefit in considering situations where line-haul capacity has recently changed, e.g., through the adoption of computer-assisted dispatching systems or the addition of track. In such situations, there are few historical data to base decisions on how to adjust the targets. A tool could help by adjusting the targets based on its predictions for delay.

3. METHODOLOGY

A comparative method of simulation-based analysis is used to compare the performance of the LDM/TTG method with traditional target time generation methods. Estimates of line-haul performance are derived by simulating line operations for a given set of targets. Initially, we simulate the performance of Base Case targets which are derived from operating data collected by a major Class I North American railroad. Next, we simulate the performance of the TTG-adjusted targets, which are created by using the TTG to adjust the Base Case (real-world) targets. The performance of the Base Case targets provides a benchmark to assess the TTG-adjusted targets' relative performance. The simulation model allows for a systematic study to consider the effects of varying traffic volume, departure time uncertainty, and track layout.

3.1. The simulation model

The discrete-event (rolling-time horizon) simulation provides a detailed representation of transient behavior in an optimal meet/pass planning process under realistic (time-constrained, stochastic) operating conditions. The inputs to the simulation include a description of the track (locations of sidings and lengths of track sections), the targeted times of departure and arrival at the points where trains' are scheduled, free running times between track sections, minimum headway distance between trains, switching delays when a train switches tracks, train lengths (which are used to determine if a train can fit into a track section), the distribution (mean and variance) for the starting departure time, and the meet/pass objective function tardiness weights. See Hallowell (1993) for a detailed description of the simulation.

We consider the simplified situation, where there are targets at the starting (origin) point and at the ending (destination) point on the rail line. In reality, there may be intermediate scheduled points within the line, where trains may dwell for a period of time, e.g. to pick up or drop-off cars and passengers. Further tests are required to consider intermediate scheduled points.

Each simulation run considers train movements that occur over a consecutive time period of about 3 days. Replanning occurs intermittently, at 4 h intervals, to take into account changes on the line, i.e., updates to the starting departure times. At times of replanning, the SCAN meet/pass planning algorithm is called to determine train movements over the next 4 h. The next four hours of train movements are fixed, the state of the simulation is then updated, and time advances. The meet/pass algorithm uses an 8 h planning horizon. This was found to be long enough to avoid line blockage (the situation where two trains cannot proceed unless one reverses). In practice, dispatchers use a planning horizon of about four to eight hours.

The simulation model is run N times over N realizations for the departure times. We use a two-stage sampling method called the Latin Hypercube (LHC) method (McKay *et al.*, 1979) with the aim of making the experimental procedure more efficient. This is a stratified sampling method which is expected to decrease the run size and produce more uniform random samples. The divisions of the strata and number of samples are selected according to Welsh's inspection method (Law and Kelton, 1991). All problems use 100 simulation runs with 20 strata, except for one particularly large problem which uses 50 runs due to excessive compute time.

The outputs of the simulation are the sample mean and standard deviation of the performance measures for individual trains, and for aggregate groupings such as by train class. Note that statistical significance for multiple response variables would require excessive compute time for the problems we tested (in the order of several weeks). The only practical recourse is to limit the sample size, but be aware that several of the resulting outputs provide only an estimate for the true means and variances.

3.2. *Performance measures*

Several metrics are used to assess the performance of a given set of targets. For each train i, let CUR_i be its cumulative unrestricted running time from origin point to destination point, D_i be the initial TTD at its origin, and A_i be the initial TTA at its destination. For simulation Run $n\varepsilon$ $\{1,2,3,....,N\}$, let $d^n{}_i$ be the actual departure time and $a_i{}^n$ be the actual arrival time. For each Train i, the outputs of Run n are defined as follows:

$$\begin{aligned} \text{departure time error at origin point} &= d^n{}_i - D_\mathrm{i} \\ \text{arrival time error at destination point} &= a^n{}_i - A_i \\ \text{delay time from origin to destination} &= a^n{}_i - d^n{}_i - CUR_i, \text{ and} \\ \text{slack time from origin point to destination point} &= A_i - D_i - CUR_i. \end{aligned}$$

Arrival error (mean and standard deviation) is the key performance measure, as it indicates on-time arrival performance with respect to the targets. Departure time error is the initial (input) cause of uncertainty in the system. The TTG must compensate for the departure time uncertainty if it is to prescribe achievable TTAs. Owing to the variability of departures, the meets and passes will be different for different simulation runs. Delay, a direct result of meets and passes, will therefore be different for each simulation run. Thus, delay is a good indicator of the complexity inherent in the system and a good target time generator should account for the variability of delay. In addition, the targets are used by the SCAN meet/pass planning algorithm to optimally resolve meet and passes. Hence, delay and arrival time error indicate how well the targets work in conjunction with an optimal meet/pass planning algorithm to improve on-time performance and decrease delay.

Slack equals the initial target arrival time minus the initial target departure time minus the free running time. The empirical work will test if a given set of targets have enough slack to account for meet/pass and departure delay.

Simulation CPU time is a surrogate variable for the complexity of the meet/pass planning process. In our tests, the complexity of meet/pass planning is dependent upon the TTDs and TTAs, since the meet/pass algorithm tries to achieve minimum tardiness cost performance. Hence, the simulation CPU times indicate the computational requirements of implementing the targets. In addition, the compute time for the TTG is an indicator of the practically of the TTG. Note that reported CPU time does not include input/output processing. The simulation and TTG are implemented on an HP-Apollo 425E workstation.

3.3. *Problem instances*

The problems represent realistic but hypothetical operating scenarios, based on historical rail operations data and interviews with personnel at several railroads (Hallowell *et al*., 1991). The analysis considers two typical rail lines from a major Class I railroad.

There are 18 problem instances which consist of the two lines, each with three traffic volumes (low, medium, high) and three scenarios for the departure distributions (low uncertainty, average uncertainty, and high uncertainty). The first rail line is 276 miles long with less than 3% double-track and 21 meet/pass points, and has three traffic scenarios: 22, 29, and 33 trains per day. The second line is of length 230 miles with a substantial amount of double-track (55%) and 39 meet/pass points, and has three traffic scenarios: 22, 26, and 28 trains per day. The traffic volume is increased by adding trains in different classes proportional to the original volume. Three days of train movements are considered in each simulation run, but only results for the second (middle) day are reported to avoid end-effects caused by truncation of the simulation planning horizon and provide for a warm start. See Hallowell (1993) for a description of the targets and track layout.

Trains are assigned one of four classes: Amtrak (passenger), Intermodal, Bulk (either empty or loaded trains carrying coal, grain, potash, etc.), or Mixed Freight. Trains within a class share similar properties such as transit times and objective function weights. The weights are as follows: 1000 for Amtrak, 266 for Intermodal, 182 for Bulk, and 163 for Mixed Freight; these values are provided by Smith (1990b).

To represent unforeseen delays that occur to trains before they depart from their origin point within the rail line, the simulation periodical revises the targeted departure times. We assume that each train's actual TTD is uniformly distributed about its initial TTD, and the TTD will be revised at most three times. During execution of the simulation each train's TTD is updated approximately 16, 10, and 4 h before its actual departure. We assume that Amtrak trains have a zero probability of being ready to depart before their TTD, and that all other trains have a 50% probability of departing early.

To facilitate a systematic analysis of uncertainty, we utilize the scalar parameter ρ, which is defined as the percentage of the standard deviation in the average departure time uncertainty scenario. The distributions for the average scenario ($\rho = 100\%$) are used to construct other scenarios by multiplying the TTD standard deviation by ρ. Table 1 presents the standard deviation for each of the three departure time uncertainty scenarios, i.e. $\rho =$ 50, 100 and 150%. Within each scenario there are three standard deviations corresponding to three TTD revisions. For instance, an intermodal train's departure time is uniformly distributed over the interval $[-30, 30]$ for the first time the TTD is revised in the average reliability scenario (SD is 17.32 min). Note that each train's TTD standard deviation will decrease as the train approaches the time of departure to represent the idea that uncertainty decreases over time.

Meet/pass interference delays are dependent upon the length of the train, the switching system and deceleration and acceleration delays that arise when a train switches from the main line to the side track. We assume a constant 10-min following headway and a constant 5-min interference delay which is assigned to a train when it takes the side track. Our analysis assumes that trains can meet and pass at any point along the line. In reality, there may be constraints on the set of feasible meet/pass points, e.g. if a train cannot stop on a steep gradient.

The final input for the analysis is the unrestricted (free running) transit times through individual track sections. They are calculated by the railroad's train performance simulation model. These times take into account such factors as locomotive power, track gradient and curvature, and air resistance. Further work is required to model variable free running times.

4. RESULTS

This section presents the results aimed at testing the performance of the TTG-adjusted targets relative to the performance of a set of corresponding Base Case targets. The difference between the Base Case and TTG-adjusted targets is their slack. The Base Case targets represent present-day target generation methods and they have a substantial amount of slack. In practice, railroads add slack to account for several sources of variability not considered in this study, such as variable free running travel times, track outages, or slow orders for maintenance. The TTG-adjusted targets are derived from the Base Case targets. The LDM/TTG attempts to assign slack in a way that makes the targets achievable under an optimal dispatch policy. For the problems tested, the TTG-adjusted targets are significantly 'tighter' (have less slack) than the Base Case targets.

Table 1. Standard deviations for the target time of departure (TTD) revisions (min per train)

Scenario (Departure uncertainty)	$\rho = 50\%$			$\rho = 100\%$			$\rho; = 150\%$		
Revision	First	Second	Third	First	Second	Third	First	Second	Third
Amtrak	8.66	6.50	4.33	17.32	12.99	8.66	25.98	19.50	12.99
Intermodal	8.66	6.50	4.33	17.32	12.99	8.66	25.98	19.50	12.99
Bulk	17.33	11.55	5.78	34.65	23.10	11.55	51.98	34.65	17.33
Mixed Freight	11.55	8.66	5.78	23.10	17.33	11.55	34.65	26.00	17.33

Nine scenarios for each of the two rail lines are constructed to test the sensitivity of performance to changes in traffic volume and departure time uncertainty. Section 4.1 reports the performance of the Base Case targets. Section 4.2 reports the performance of the TTG-adjusted targets. In Section 4.3, we compare and contrast the performance for the two sets of targeted times. The results are summarized by overall averages owing to the magnitude of the data sets. The reader is referred to Hallowell (1993) for more detailed results by train class and by individual train.

4.1. Performance of the Base Case targets

Overall average performance under the Base Case targeted times for Lines I and II is presented in Tables 2 and 3, respectively. The column headings designate the traffic volume and the departure time uncertainty scenario. The tables present mean and standard deviation (SD) of arrival error, slack, cumulative unrestricted running time (CUR), SD departure time error, mean and SD of delay, total simulation CPU time, the number of simulation runs, and the number of trains per day.

4.1.1. On-time performance. Mean arrival error is a key performance measure as it indicates the accuracy with which the targeted times are achieved. In the medium volume scenario with $\rho = 100\%$, trains arrive, on average, 34.49-min early for Line I, and 85.62-min early for Line II. It is generally the lower priority trains, i.e. Bulk and Mixed Freight, that arrive very early, while Amtrak trains arrive about on-time, and Intermodal trains arrive about 30-min early. Given that the average transit times are about 5–7 h, the large negative mean arrival errors suggest that the Base Case targets are very 'loose' for the majority of trains under the optimal meet/pass planning policy we modeled.

Arrival time is dependent upon departure time variance. Mean arrival error is fairly insensitive to changes in departure time uncertainty, compared with the SD of arrival error, which is very sensitive to this uncertainty. For instance, in the Line I/medium volume scenario, as departure uncertainty increases from $\rho = 50\%$ to $\rho = 150\%$, overall mean error increases about 11% (from −34.22 to −30.49 min), while overall SD error increases about 27% (from 17.88 to 22.65 min).

Table 2. Overall average performance per train, Base Case targets, Line I

Volume	Low			Medium			High		
ρ Departure uncertainty	50% (Low)	100% (Avge)	150% (High)	50% (Low)	100% (Avge)	150% (High)	50% (Low)	100% (Avge)	150% (High)
Mean arrival error (min)	−61.99	−59.64	−58.71	−34.32	−34.49	−30.49	−29.69	−27.75	−26.83
SD arrival error (min)	12.87	18.08	23.69	17.88	19.45	22.65	15.62	19.10	22.41
Slack (min)	113.00	113.00	113.00	108.00	108.00	108.00	119.00	119.00	119.00
CUR (min)	290.59	290.59	290.59	301.09	301.09	301.09	301.97	301.97	301.97
SD departure error (min)	2.56	5.12	7.67	2.56	5.12	7.69	2.58	5.16	7.74
Mean delay (min)	49.71	51.23	51.11	72.10	71.01	74.08	89.54	89.72	89.72
SD delay (min)	12.62	17.86	22.30	18.04	19.76	23.21	16.32	19.75	24.01
Simulation CPU time (min)	11	11	13	80	99	145	447	1235	1015
No. simulation runs	100	100	100	100	100	100	50	50	50
No. trains per day	22	22	22	28	28	28	33	33	33

Table 3. Overall average performance per train, Base Case targets, Line II

Volume	Low			Medium			High		
ρ Departure uncertainty	50% (Low)	100% (Avge)	150% (High)	50% (Low)	100% (Avge)	150% (High)	50% (Low)	100% (Avge)	150% (High)
Mean arrival error (min)	−88.09	−87.80	−85.03	−86.27	−85.62	−84.19	−82.53	−81.01	−79.35
SD arrival error (min)	13.06	16.23	20.27	15.74	17.64	20.89	15.33	18.06	22.54
Slack (min)	117.00	117.00	117.00	123.00	123.00	123.00	122.00	122.00	122.00
CUR (min)	362.27	362.27	362.27	365.31	365.31	365.31	366.04	366.04	366.04
SD departure error (min)	2.46	4.92	7.38	2.47	4.94	7.41	2.47	4.95	7.42
Mean delay (min)	28.36	28.29	30.23	36.32	36.67	37.30	38.93	40.16	41.16
SD delay (min)	12.09	14.13	16.16	15.11	16.22	18.62	14.96	17.06	19.97
Simulation CPU time (min)	21	27	31	66	91	231	34	48	64
No. simulation runs	100	100	100	100	100	100	100	100	100
No. trains per day	22	22	22	26	26	26	28	28	28

A reliable set of targets must compensate for the variability in meets and passes. The overall SD of arrival error is greater than the SD of delay in 12 out of the 18 scenarios. The six scenarios where the overall SD of arrival error is lower than the overall SD of delay are the medium and high volumes for Line I. Generally speaking, the difference between these six scenarios and the others is that these six scenarios have a significantly greater proportion of trains which arrive tardy. In these scenarios, trains arrive, on average, about 30 min early, but frequently arrive tardy, compared with the other twelve scenarios, where trains arrive on average more than twice as early and are rarely tardy. We hypothesize that our minimum tardiness cost meet/pass algorithm makes less arbitrary meet/pass decisions when the targets are tighter. Hence, the SD of arrival error is lower than the SD of delay for those six scenarios which have the tightest on-time constraints.

4.1.2. Slack. Early arrivals are a result of excess slack. Average slack is about 38% of the average cumulative unrestricted running time in Line I and 33% in Line II. Note that the method of Base Case targeted time construction did not attempt to alter the amount of slack to account for varying degrees of uncertainty or congestion. As a result, average slack per train is similar for different volume scenarios.

4.1.3. Delay. Enough slack needs to be added to the targeted transit time to account for meet/pass delay. Slack is significantly larger than mean delay for most trains, even in the high volume scenarios and, therefore, trains arrive early. Mean delay is fairly insensitive to changes in departure time variability but is very sensitive to changes in traffic volume. For instance, in the Line I/medium volume scenario, as departure uncertainty increases from $\rho = 50\%$ to $\rho = 150\%$, overall mean delay increases from 72.10 to 74.08 min. In comparison, when ρ remains constant at 100% and traffic increases from low to high, mean delay increases substantially from 51.23 to 89.72 min. This is expected since increases in traffic volume result in increases in meet/pass delay.

The major challenge in determining achievable TTAs stems from the variability in delay. In the Line I/medium volume scenario, as the average SD of the TTDs increases from 2.56 to 7.69 min, the average SD of delay increases from 18.04 to 23.21 min. The variance of delay is also sensitive to congestion. In Line II with $\rho = 100\%$, as volume increases from low to medium to high, average SD delay increases from 14.13 to 16.22 to 17.06 min. Generally speaking, increases in congestion lead to increases in the variance of delay, but we report several cases where this did not occur (compare the SD of delay for the medium and high volumes). We expect that low overall SD of delay for the high volume scenarios relative to the medium volume scenarios is due to the length of the simulation planning horizon. The high volume scenarios represent about 60 h of simulated train movements, while the medium and low volume scenarios represent about 84 h. As a result, we expect that the variability in the system caused by the propagation of meet/pass delay may be less for the high volume problems. Nevertheless, for most trains the SD of delay is 2–4-times the departure time SD which is input to the system.

4.1.4. Computational performance. The compute time of the simulation is substantially affected by traffic volume, and to a lesser extent it is also affected by uncertainty. In Line I with $\rho = 100\%$, as the volume increases from low to high, the CPU time rose from 11 to 1234 min. In comparison, as departure time uncertainty increases from $\rho = 50\%$ to $\rho = 150\%$ for the Line I/medium volume scenario, the CPU time increases moderately from 80 to 145 min. Note that in the medium and high volume scenarios for Line I the amount of time used by the meet/pass planning algorithm is constrained in order to represent a real-world (time-constrained) meet/pass planning process. Thus, congestion significantly effects the computational complexity of the meet/pass planning process, even with the loose, easy to achieve, Base Case targets.

4.2. Performance of the TTG-adjusted targets

This section considers the performance of the TTG-adjusted targets. The results are given in Tables 4 and 5 for Lines I and II, respectively. These tables contain the same measures as reported earlier for the Base Case targets. In addition, the TTG compute time is reported in Row 10.

4.2.1. On-time performance. Low overall mean arrival time errors indicate that trains arrive near their targets, although there is some variability in on-time performance for individual trains.

Overall average arrival error tends to be insensitive to changes in departure time uncertainty. For instance, in Line I/medium volume scenario, mean arrival error is -0.57 min for $\rho = 50\%$, -2.22 min for $\rho = 100\%$, and -3.82 min for $\rho = 150\%$. Detailed analysis indicates that more than 76% of the absolute mean arrival errors by train class are less than 10 min for both lines, and the percentage of individual trains that arrive within ± 10, ± 20, and ± 30 min of the TTAs are, respectively, 59, 82, and 90% for Line I and 58, 79, and 89% for Line II.

When absolute arrival error is large (say greater than 20 min), it more frequently represents earliness, rather than lateness. However, one situation where the TTG consistently assigns very tight targets is for the Amtrak class in Line II's medium volume scenario, e.g. where $\rho = 100\%$ Amtrak trains are, on average, 20-min late. Considering the high degree of stochasticity in the system and that the average transit time is greater than 6 h, the fairly low arrival time errors indicate that the TTG is relatively effective at determining targets that are achieved with 50% percentile reliability.

The variability of arrival time error is an indicator of the reliability of the TTAs. The SD of arrival error is only slightly sensitive to changes in congestion. For instance, in Line I with $\rho = 100\%$, overall SD of arrival error is 14.43 min for the low volume, 16.16 min for the medium volume, and 11.79 min for the high volume. We expect the decrease in standard deviation from the medium volume to the high volume is due to the fact that we simulated 3 days train movements for the medium traffic volume case, compared with 2.5 days train movements for the high volume case, i.e. there is less propagation of delay for the high volume scenario. Generally speaking, the data suggest that the TTG provides targets that achieve fairly consistent variance of on-time arrival performance under changes to traffic volume.

The SD of arrival error also indicates how well the TTG's targets work in conjunction with an optimal meet/pass planning algorithm to absorb the variability in the system. The SD of arrival times typically increase as the SD of the departure variability increases. For the Line I/medium volume scenario, as the average SD of the departures increases from 2.56 to 5.12 to 7.69 min, overall SD of arrival errors increases from 13.86 to 16.16 to 18.07 min. However, for the Line II/high volume scenario, as the average SD of the departure times increases from 2.56 to 5.12 to 7.69 min, the overall SD of arrival error decreases from 10.85 to 9.52 to 8.24 min. Furthermore, the variability of meet/pass delay is sometimes greater than the variability of arrival times. The overall SD of arrival error is less than the overall SD of delay in 44% of the scenarios, and it is nearly equal to (+ 2 min) the SD of delay in over 89% of the scenarios. This suggests that the TTG derives targets that have the potential to reduce the variability in the system under an optimal dispatching policy.

4.2.2. Slack. To ensure that the trains' TTAs are reliable, the TTG adds slack to compensate for train delay. In all cases, the average slack per train increases as traffic volume increases. In Line I with $\rho = 100\%$, as volume increases from low to medium to high, overall average slack increases from 32 to 56 to 67 min. Another reason that the TTG adds slack is to hedge against uncertainty in the departure times. In Line I for the medium volume, as uncertainty increases from $\rho = 50\%$ to $\rho = 100\%$ to $\rho = 150\%$, average slack increases from 51 to 56 to 60 min. In addition, note that Line II has less slack than Line I, even when slack is normalized by such relevant factors as free running times or trip miles. For Line I the average slack is 11.01, 18.6 and 22.2% of the CUR for the low, medium, and high volumes, respectively; for Line II, slack as a percentage of CUR is 5.8, 6.8 and 7.4% for the low, medium, and high volumes, respectively. This seems reasonable since Line I is only about 3% double-track, while Line II is over 55% double-track. The TTG derived values for slack confirm our intuitive understanding of how train delay is affected by traffic volume, departure time reliability, and line capacity as measured by percent double-track.

4.2.3. Delay. Mean meet/pass delay is moderately sensitive to traffic volume, but it is relatively insensitive to departure uncertainty. As expected, the addition of trains results in increases in average delay. For instance, in Line I with $\rho = 100\%$, as traffic volume increases from low to medium to high, average mean delay increases from 33 to 52 to 56 min. Increases in delay also result from increases in departure uncertainty, usually only several minutes, although in Line II there are a few cases where overall mean delay decreases a little as uncertainty increases. The results imply that the TTG assigns enough slack to account for meet/pass delay under varying conditions.

Table 4. Overall average performance per train, TTG-adjusted targets, Line I

Volume	Low			Medium			High		
ρ Departure uncertainty	50% (Low)	100% (Avge)	150% (High)	50% (Low)	100% (Avge)	150% (High)	50% (Low)	100% (Avge)	150% (High)
Mean arrival error (min)	0.31	3.40	0.86	−0.57	−2.22	−3.82	−14.30	−11.74	−15.50
SD arrival error (min)	9.51	14.43	16.94	13.86	16.16	18.07	12.22	11.79	13.08
Slack (min)	34.00	32.00	37.00	51.00	56.00	60.00	66.00	67.00	71.00
CUR (min)	290.59	290.59	290.59	301.09	301.09	301.09	301.97	301.97	301.97
SD departure error (min)	2.56	5.12	7.67	2.56	5.12	7.69	2.58	5.16	7.74
Mean delay (min)	32.95	33.44	34.53	49.40	51.72	53.00	51.66	55.64	55.77
SD delay (min)	10.11	15.02	17.02	14.30	16.70	19.05	12.16	11.70	12.93
Simulation CPU time (min)	492	743	795	3001	1155	2415	3731	2474	2267
No. simulation runs	100	100	100	100	100	100	50	50	50
TTG CPU time (min)	67	20	37	18	9	22	246	342	209
No. trains per day	22	22	22	28	28	28	33	33	33

Table 5. Overall average performance per train, TTG-adjusted targets, Line II

Volume	Low			Medium			High		
ρ Departure uncertainty	50% (Low)	100% (Avge)	150% (High)	50% (Low)	100% (Avge)	150% (High)	50% (Low)	100% (Avge)	150% (High)
Mean arrival error (min)	1.57	0.61	−0.72	3.75	0.42	1.32	1.45	−0.62	−1.02
SD arrival error (min)	10.12	12.75	16.70	12.30	15.00	18.24	10.85	9.52	8.24
Slack (min)	20.00	21.00	22.00	22.00	25.00	26.00	27.00	27.00	28.00
CUR (min)	362.27	362.27	362.27	365.31	365.31	365.31	366.04	366.04	366.04
SD departure error (min)	2.46	4.92	7.38	2.47	4.94	7.41	2.47	4.95	7.42
Mean delay (min)	21.28	20.60	19.61	25.50	24.92	25.95	28.14	26.43	26.93
SD delay (min)	8.91	10.14	12.08	12.08	13.53	15.19	10.89	9.46	8.59
Simulation CPU time (min)	1676	383	2498	1451	999	4761	2821	3007	1989
No. simulation runs	100	100	100	100	100	100	100	100	100
TTG CPU time (min)	20	53	51	93	58	82	100	38	48
No. trains per day	22	22	22	26	26	26	28	28	28

If the TTG is to be effective at deriving achievable TTAs, then it must compensate for the variability in meet/pass delay which is caused by departure time variability. In the Line I/ medium volume scenario, as the average SD of the departure times increases from 2.56 to 5.12 to 7.69 min, the average SD of delay increases from 14.30 to 16.70 to 19.05 min. As expected, the variance of delay for Amtrak and Intermodal trains is lower than for Bulk and Mixed Freight trains. One possible reason that Amtrak and Intermodal trains have a lower SD of delay is that these trains have the lower departure time variance. Another reason is that Amtrak and Intermodal trains have a higher weight (cost) for tardiness in the dispatching objective and, therefore, they are less likely to be delayed.

4.2.4. Computational performance. The simulation's CPU times serves as a surrogate variable that indicates the complexity of the meet/pass planning process. CPU time is highly sensitive to congestion. In Line I with $\rho = 100\%$, as traffic volume increases from low to medium to high, CPU time increases dramatically from 743 to 1155 to 2474 min. While higher congestion usually causes higher simulation CPU times, the effect of departure uncertainty on CPU times is fairly random. For instance, for Line I with the medium volume, CPU time is 3001 min for $\rho = 50\%$, 1155 min for $\rho = 100\%$, and 2415 min for $\rho = 150\%$. We expect that the large simulation compute times are due to two reasons. First, it is difficult to find good (low tardiness cost) initial meet/pass planning solutions when re-planning occurs. Second, it is practically impossible to find optimal (minimum tardiness cost) meet/pass solutions given the very tight (TTG-derived) targets.

The TTG CPU time indicates the practicality of the TTG. The average TTG compute time is 107 min for Line I and 60 min for Line II, although there are several cases when it is substantially higher than average (e.g. Line I for the high volume). The time it takes the TTG to adjust the targets does not appear to be strongly correlated to any particular factor, although it tends to be larger for the high traffic volume scenarios. We expect that the TTG compute times would be adequate for planning purposes, and with a faster computer the TTG might also be fast enough for real-time target updating.

The TTG's compute times are substantially lower than the simulation compute times for all the scenarios we tested. As an example, for Line I with the medium volume and $\rho = 100\%$, the total simulation time is 1155 min and the TTG time is 9 min. Similar results are found for Line II, with the medium volume and $\rho = 100\%$: the simulation time is 990 min and the TTG time is 58 min. The data suggest that the TTG provides a more practical approach than our Monte Carlo simulation for adjusting the targets.

4.3. *Comparing the performance of the TTG-adjusted targets to the base case targets*

The remainder of this section examines the TTG targets performance relative to the Base Case targets. In what follows, we refer to Tables 6 and 7 which summarize the results for Lines I and II. These tables contain the overall average results for the Base Case and the TTG targets.

4.3.1. *On-time performance.*

Mean arrival error indicates if the targeted arrival times are achievable. Large negative arrival time error implies that trains generally arrive very early for the Base Case, while small errors imply that trains arrive near their targeted times for the TTG-adjusted targets. The Base Case has only 19% of the mean arrival errors per train less than 10 min

Table 6. Comparison of overall average performance for Line I

Volume		Low			Medium			High		
ρ Departure uncertainty		50% (Low)	100% (Avge)	150% (High)	50% (Low)	100% (Avge)	150% (High)	50% (Low)	100% (Avge)	150% (High)
Mean arrival error	BC	−61.99	−59.64	−58.71	−34.32	−34.49	−30.49	−29.69	−27.25	−26.83
per train (min)	TTG	0.31	3.40	0.86	−0.57	−2.22	−3.82	−14.30	−11.74	−15.50
SD arrival error	BC	12.87	18.08	23.69	17.88	19.45	22.65	15.62	19.10	22.41
per train (min)	TTG	9.51	14.43	16.94	13.86	16.16	18.07	12.22	11.79	13.08
Slack	BC	113.00	113.00	113.00	108.00	108.00	108.00	119.00	119.00	119.00
per train (min.)	TTG	34.00	32.00	37.00	51.00	56.00	60.00	66.00	67.00	71.00
Mean CUR time	BC	290.59	290.59	290.59	301.09	301.09	301.09	301.97	301.97	301.97
per train (min)	TTG	290.59	290.59	290.59	301.09	301.09	301.09	301.97	301.97	301.97
SD departure error	BC	2.56	5.12	7.67	2.56	5.12	7.69	2.58	5.16	7.74
per train (min)	TTG	2.56	5.12	7.67	2.56	5.12	7.69	2.58	5.16	7.74
Mean delay	BC	49.71	51.23	51.11	72.10	71.01	74.08	89.54	89.72	89.72
per train (min)	TTG	32.95	33.44	34.53	49.40	51.72	53.00	51.66	55.64	55.77
SD delay	BC	12.62	17.86	22.30	18.04	19.76	23.21	16.32	19.75	24.01
per train (min)	TTG	10.11	15.02	17.02	14.30	16.70	19.05	12.16	11.70	12.93
Simulation CPU	BC	0.11	0.11	0.13	0.80	0.99	1.45	7.36	24.69	13.24
time per run (min)	TTG	4.92	7.43	7.95	30.01	11.55	24.15	74.62	49.48	45.33

Table 7. Comparison of overall average performance for Line II

Volume		Low			Medium			High		
ρ Departure uncertainty		50 (Low)	100 (Avge)	150 (High)	50 (Low)	100 (Avge)	150 (High)	50 (Low)	100 (Avge)	150 (High)
Mean arrival error	BC	−88.09	−87.80	−85.03	−86.27	−85.62	−84.19	−82.53	−81.01	−79.35
per train (min)	TTG	1.57	0.61	−0.72	3.75	0.42	1.32	1.45	−0.62	−1.02
SD arrival error	BC	13.06	16.23	20.27	15.74	17.64	20.89	15.33	18.06	22.54
per train (min)	TTG	10.12	12.75	16.70	12.30	15.00	18.24	10.85	9.52	8.24
Slack	BC	117.00	117.00	117.00	123.00	123.00	123.00	122.00	122.00	122.00
per train (min)	TTG	20.00	21.00	22.00	22.00	25.00	26.00	27.00	27.00	28.00
Mean CUR time	BC	362.27	362.27	362.27	365.31	365.31	365.31	366.04	366.04	366.04
per train (min)	TTG	362.27	362.27	362.27	365.31	365.31	365.31	366.04	366.04	366.04
SD departure error	BC	2.46	4.92	7.38	2.47	4.94	7.41	2.47	4.95	7.42
per train (min)	TTG	2.46	4.92	7.38	2.47	4.94	7.41	2.47	4.95	7.42
Mean delay	BC	28.36	28.29	30.23	36.32	36.67	37.30	38.93	40.16	41.16
per train (min)	TTG	21.28	20.60	19.61	25.50	24.92	25.95	28.14	26.43	26.93
SD delay	BC	12.09	14.13	16.16	15.11	16.22	18.62	14.96	17.06	19.97
per train (min)	TTG	8.91	10.14	12.08	12.08	13.53	15.19	10.89	9.46	8.59
Simulation CPU	BC	0.21	0.27	0.31	0.66	0.91	2.31	0.34	0.48	0.64
time per run (min)	TTG	16.76	3.83	24.98	14.51	9.99	47.61	28.21	30.07	19.89

for Line I and 11% for Line II. In comparison, the TTG has 78% of the mean errors per train less than 10 min for Line I, and 74% for Line II. Although on-time performance varies by individual trains, when the TTG's reported arrival error is large (e.g. greater than 20 min) then it typically represents earliness, rather than lateness.

The reliability of on-time performance is measured by the SD of arrival error. The overall average SD of arrival error is lower for the TTG in all of the 18 scenarios with a typical reduction of 20%. For the Base Case, only 33% of the average SD of arrival error per train are less than 15 min for Line I and 41% for Line II. The corresponding percentiles for the TTG are 58% for Line I and 85% for Line II. Overall, the reliability of on-time performance, as measured by the variance of arrival times is considerably improved by the TTG.

Typical arrival time error distributions are illustrated in Figs 3 and 4 for the Base Case and the TTG, respectively. These histograms are constructed from 2200 samples (100 simulation runs with 22 trains) for Line I with the low volume and $\rho = 100\%$. The TTG histogram illustrates that the TTG's arrival time error is roughly normal and almost always contained in the interval ± 40 min. Although there is a high variability of on-time performance for both the Base Case and the TTG, the Base Case arrival errors are significantly more dispersed. On the other hand, trains more frequently arrive late for the TTG-derived targets.

4.3.2. Slack. The difference between the Base Case and TTG-adjusted targets is slack. In Line I the percent reduction in average slack relative to the Base Case is about 70, 48 and 43% for the low, medium and high volume scenarios, respectively. For Line II, the corresponding percent reductions are 82, 80 and 78%. Since the percent reduction in slack decreases as volume increases, the effects of tightening the targets are emphasized in the low volume scenarios.

To determine reliable TTAs, enough slack needs to be added to the free running times to account for meet/pass delay. The TTG-adjusted targets represent the situation where the targeted times are tight, since the TTG goal is to assign enough slack to cover mean delay. The comparison of average slack to mean delay indicates that the Base Case (real-world) targets have substantially more slack than mean delay, while the TTG targets have roughly enough slack to cover mean delay.

We expect that the TTG-adjusted targets are tighter, in part, because the Base Case targets are based on an (inefficient) fixed-priority dispatch policy, while the TTG-adjusted targets are based on a (more efficient) minimum tardiness dispatch policy. In particular, note that lower priority

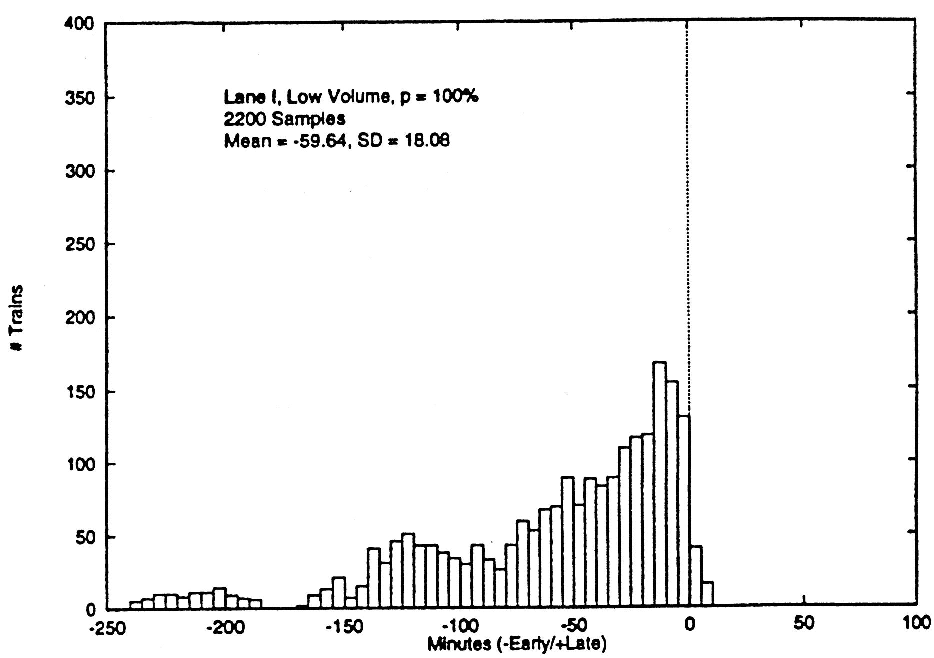

Fig. 3. Arrival time error histogram, Base Case targets.

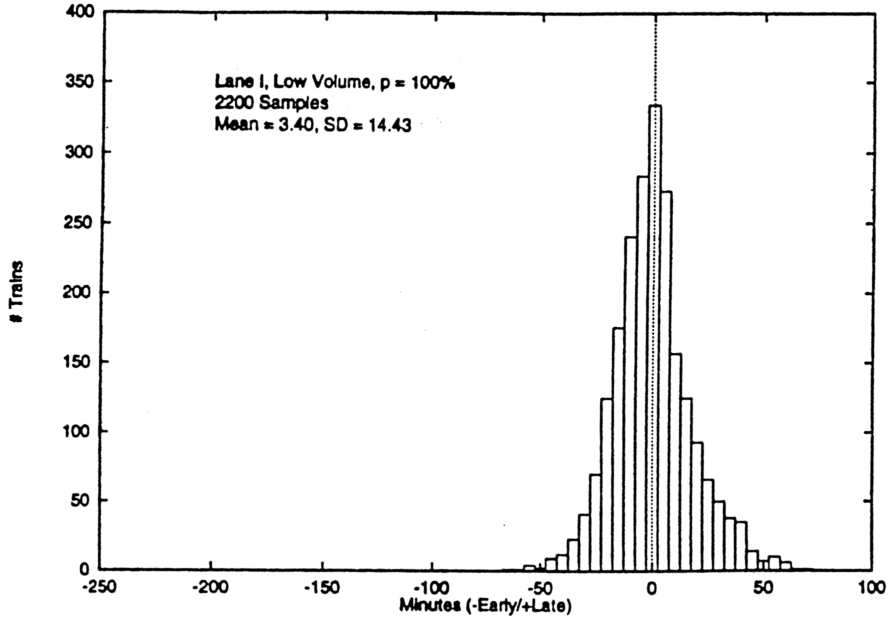

Fig. 4. Arrival time error histogram, TTG-adjusted targets.

Bulk and Mixed Freight trains have a lot more slack in the Base Case than in the TTG-adjusted targets, while Amtrak trains have slightly less slack in the Base Case. We expect that the Base Case targets incorporate a lot more slack for the lower priority trains, since these trains are delayed frequently under real-world fixed-priority dispatching.

4.3.3. Delay. The tighter TTG-adjusted targets, when used in conjunction with an optimal meet/pass planning policy, achieve fairly significant reductions in delay. The range of percent reduction in mean delay is 27 to 42% for Line I, and −25% to −35% for Line II. The range of percent reduction in SD delay is 15 to 46% for Line I, and 17% to 57% for Line II.

The data suggest that overall reductions in delay are a result of the more precise targets derived by the TTG. The targets are an important input to the simulated system since the meet/pass planning algorithm tries to resolve meets and passes in a way that minimizes tardiness costs. Since this objective says nothing about the cost of delay, the algorithm is indifferent between two plans with equal tardiness costs, even if one of the plans entails much greater delay. Therefore, the meet/pass algorithm assigns delay in a less arbitrary manner when the targets are tighter. As a result, mean and SD of delay are lower for the tighter (TTG) targets than for the looser (Base Case) targets, but this is at the expense of increased tardiness.

4.3.4. Computational performance. The simulation CPU time is greater for the TTG-adjusted targets than for the Base Case targets. For Line I, the largest percent increase in CPU time occurs in the low volume scenario with $\rho = 100\%$ (CPU time per run is 0.11 min for the Base Case and 7.43 min for the TTG). For Line II, the largest percent increase in CPU time occurs in the high volume scenario with $\rho = 50\%$ (CPU time per run is 0.34 min for the Base Case and 28.21 min for the TTG). We expect the meet/pass planning algorithm within the simulation required more compute time for the TTG-adjusted targets because it is more difficult to determine minimum cost plans under tight targets.

5. SUMMARY

In summary, this paper presents a simulation-based analysis to examine the application and validation of the LDM/TTG under an optimal, on-time performance-based dispatching policy. The authors' primary objective is to validate that the LDM/TTG prescribes achievable targets. A second objective is to investigate the potential impacts of adopting a train scheduling methodology aimed at tightening the schedules.

5.1. Conclusions

For the problems tested, the LDM/TTG prescribes targets that are reasonably achievable over the range of stochastic operating conditions considered. Approximately 59% of the trains arrive within 10 min of their TTA and 87% arrive within 30 min. The average SD of arrival time error, an indicator of the reliability of the TTAs, is about 15 min and generally does not exceed 20 min. The difficulty in determining achievable TTAs stems from the variability in departure times and meets/pass delays which must be compensated for by a good target time generating algorithm. In the average scenario for our analysis, trains depart ± 17 min about their targeted time of departure and the SD of delay is 15 min. Considering the degree of variability in the system, we conclude that the LDM/TTG is moderately effective at determining achievable TTAs, as measured by arrival time error.

According to the comparative analysis, based on the examples used, there are opportunities to improve the productivity of line operations without sacrificing reliability if the LDM/TTG is used in the scheduling process. The LDM/TTG may be used as a 'rough-cut', quick and interactive tool during the scheduling process, with a more detailed simulation used for 'fine-tuning'.

The results suggest that traditional rail scheduling methodology may result in poor utilization of line-haul capacity (both track and rolling stock), especially with availability of optimal train dispatching. The majority of trains in the Base Case arrive very early and, therefore, have a substantial amount of excess slack. The only trains which had tight targets are high priority trains, i.e., passenger trains, and they are rarely delayed in real-world operations. Our findings suggest that such schedules, with optimal dispatching, encourage the dispatcher (or dispatching algorithm) to give high priority to a few trains. Consequently, the majority of trains incur a substantial amount of delay as they wait for the few high priority trains to pass. When the meet/pass planning goal is to minimize tardiness, the dispatching logic is indifferent between two plans with the same tardiness, even if one plan has much less delay. As a result, the loose Base Case targets produced relatively high mean delay, as well as high variability of arrival times.

In theory, some excess slack could be used wisely by slowing down or pacing trains (see Kraay *et al.*, 1991 for the development of models and tools for the optimal pacing problem). Pacing trains, as opposed to running trains at full speed, is expected to reduce fuel and maintenance costs and reduce the variability of train arrivals. Our analysis assumed that trains are allowed to arrive before their target arrival time. If the simulation incorporated pacing we expect the arrival time error mean and variance would have decreased, although average delay would have increased.

For the problems we considered, substantial reductions in train delay were achieved by the LDM/TTG in conjunction with an optimal meet/pass algorithm. A conservative estimation of the gains is a 20% reduction in delay without increased lateness. This improvement is achieved by the more precise targets prescribed by the LDM/TTG, which resulted in less arbitrary and more consistent assignment of delay.

In practice, the provision of more precise targets is expected to increase productivity and reliability of rail operations by fostering a more cost-effective allocation of resources and proactive planning. Reductions in delay may lead to decreased operating costs associated with increases in the utilization of locomotives, rolling stock, trackage, terminal resources and labor. In the long term, decreases in delay may translate into decreases in infrastructure and ownership expense.

However, the most significant economic benefits from improved target generation are expected to be derived from reductions in shipment transit times and from increased reliability of shipment times since this enhances the marketability of services. Even a fairly modest improvement in the reliability of shipment delivery time is expected to yield sizable returns (Allen *et al.*, 1985; Hertenstein and Kaplan, 1991; Moody, 1993). Hence, increases in price or in market share may provide sole justification for the adoption of a system-wide scheduling and operating strategy aimed at the precision scheduling.

The execution of tighter, precision schedules requires a complementary system-wide planning methodology and operating policy. According to Harker and Ward, a dynamic real-time scheduling strategy would entail an intermediate group of employees with network-wide purview (Harker, 1989; Harker and Ward, 1991). This group periodically updates the targeted times. When disruptions to the system arise, the best corrective actions are identified and passed down through the control hierarchy, to be used as operating targets by dispatchers and other operations personnel.

Railroads would have to disrupt the traditional hierarchical operations management structure to allow for the intermediate network group. A more dynamic scheduling process transfers part of authority and burden of decisions from the local decision-makers (e.g. dispatchers) to the intermediate group which controls the flow of traffic and assigns windows for track maintenance. For many railroads there would be costs associated with the reorganization of the traditional rail management structure, in order to introduce the intermediate group.

The execution of tighter, precision schedules may also require investments in decision support and information technology. Our analysis indicates, unsurprisingly, that the computational requirements of a minimum tardiness cost meet/pass planning policy are higher when target/scheduled arrival times are tightened. Hence, it is critical to ensure that adequate decision-support technology is deployed to meet the computational demands imposed by tightening the targeted times. Precision execution of the schedules also requires information flow to pass the goals to operations personnel, and to update the status of operations and progress towards the plan. Information infrastructure to enable such data flows would include database and communication technology.

5.2. Future research

A promising application for the LDM/TTG is to automate the manually intensive schedule adjustment process. This paper examined using the TTG under on-time performance-based dispatching. The model could also be used to adjust the schedules for the fixed-priority case in order to increase the productivity of schedulers. Further work is needed to validate that the LDM/TTG provides reliable targets under fixed-priority dispatch.

One promising area of research is to modify the LDM/TTG so that it prescribes targets which may be achieved with a desired level of schedule reliability. The tests presented in this paper indicate that the LDM/TTG provides fairly achievable targets, as measured by a low mean arrival time error. In practice, however, we expect that railroads may want to have the targets be achieved with a desired probability of non-tardiness, such as the 85th, 90th, 95th percentile. This would require a two-moment estimate of delay, as developed in Chen and Harker (1990), but work is needed to validate such a tool for scheduling.

Another area for future research is the development of a tactical network-wide train scheduling model. Such a tool is needed to determine system-wide schedules and the allocation of resources given constraints on physical resources and level of service objectives defined at the strategic level. We expect that the LDM/TTG could be incorporated within a higher-level network scheduling tool to ensure that line transit times are consistent with train performance and capacity. Formulating a computational tractable network model for interactive schedule development, one that considers the interactions between lines and yards, remains a challenge.

A final area for work is to analyze the benefits of alternative scheduling methodologies. A comprehensive study would need to link the expected operational outcomes of alternative methodologies to cost and revenue impacts. As a starting point, the research might utilize the effects for arrival time error and delay estimated herein. In addition, the choice of scheduling methodology has major implications for technology investment, for organization culture, and for organizational structure, and these would also need to be considered.

Acknowledgements—This work was supported by the National Science Foundation Presidential Young Investigator Award ECE-8552773, the Burlington Northern Railroad, and the Fishman-Davidson Center for the Study of the Service Sector at the Wharton School. The comments of Dejan Jovanović, David Kraay, Mike Smith, and Randy Resor are warmly acknowledged.

REFERENCES

Allen, W. B., Mahmoud, M. M. and McNeil, D. (1985) The importance of time in transit and reliability of transit time for shippers, receivers, and carriers. *Transportation Research* **B19**, 447–456.

Chen, B. and Harker, P. T. (1990) Two moments estimation of the delay on single-track rail lines with scheduled traffic. *Transportation Science* **24**, 261–275.

Hallowell, S. F. (1993) Optimal dispatching under uncertainty: with application to railroad scheduling. Ph. D. Thesis, Department of Systems Engineering, University of Pennsylvania, U.S.A.

Hallowell, S. F. and Harker, P. T. (1996) Predicting on-time line-haul performance in scheduled railroad operations. *Transportation Science* **30**, 364–378.

Hallowell, S. F., Harker, P. T., Kraay, D. R. and Sheen, R. (1991) A statistical analysis of schedule deviation, *Department of Operations and Information Management Research Report*, University of Pennsylvania, Philadelphia, PA.

Harker, P. T. (1989) The use of advanced train control systems in scheduling and operating railroads: models, algorithms, and applications. *Transportation Research Record* **1263**, 101–110.

Harker, P. T. (1995) Services and technology: reengineering the railroads. *Interfaces* **25**, 72–80.

Harker, P. T. and Ward, J. (1991) Management and information systems: components of a successful ATCS. *Transportation Research Record* **1314**, 21–30.

Hertenstein, J. H. and Kaplan, R. S. (1991) Burlington Northern: the ARES decision (a). Publishing Division, Harvard Business School, Boston, MA.

Jovanović, D. and Harker, P. T. (1990) A decision support system for train dispatching: an optimization-based methodology. *Journal of the Transportation Research Forum* **31**, 25–37.

Jovanović, D. and Harker, P. T. (1991) Tactical scheduling of rail operations: the SCAN I system. *Transportation Science* **25**, 46–64.

Kraay, D., Harker, P. T. and Chen, B. (1991) Optimal pacing of trains in freight railroads: model formulation and solution. *Operations Research* **39**, 82–99.

Kraay, D. and Harker, P. T. (1995) Real-time scheduling of freight railroads. *Transportation Research*-**B** **29**, 213–229.

Lang, A. S. and Martland, C. D. (1972) Reliability in railroad operations: studies in railroad operations and economics. Volume 8, Report No. R72-74, Transportation Systems Division, Massachusetts Institute of Technology, Cambridge, MA.

Law, A. M. and Kelton, W. D. (1991) *Simulation Modeling and Analysis*. McGraw-Hill, New York.

Martland, C. D. (1982) PMAKE analysis: predicting rail yard time distributions using probabilistic train connection standards. *Transportation Science* **16**, 476–506.

Martland, C. D., Little, P. and Sussman, J. M. (1993) Service management in the rail industry. Working Paper, Center for Transportation Studies, Massachusetts Institute of Technology, Cambridge, MA.

Martland, C. D. and Smith, M. E. (1990) Estimating the impact of advanced dispatching systems on terminal performance. *Journal of the Transportation Research Forum* **30**, 286–300.

McKay, M. D., Beckman, R. J. and Conover, W. J. (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239–245.

Moody, H. G. (1993) Resource paper on command, control, communications and information systems. AAR/FRA/TRB Conference on Railroad Freight Transportation Research Needs, Bethesda, MD, 12–14 July 1993.

Smith, M. E., Patel, P. K., Resor, R. R. and Kondapalli, S. (1990) Benefits of the meet/pass planning and energy management subsystems of the advanced railroad electronics systems (ARES). *Journal of the Transportation Research Forum* **30**, 301–309.

Smith, M. E. (1990a) Train delay analysis: alternative methods and implications. Research Paper, Burlington Northern Railroad, April 1990.

Smith, M. E. (1990b) Keeping trains on schedule: on-line planning systems for the advanced railroad electronics system (ARES). *Journal of the Transportation Research Forum* **31**, 17–24.