Decision Support System for Train Dispatching: An Optimization-Based Methodology

DEJAN JOVANOVIĆ AND PATRICK T. HARKER

The authors argue that the primary purpose of dispatching tools is to allow trains to arrive on time rather than to minimize train delays in general, and present a new methodological framework for the role of computer-aided train-dispatching (CAD) systems. One of the biggest obstacles to the successful implementation of an optimal CAD system is the combinatorial nature of the optimal train-dispatching problem and the need for optimization algorithms that could provide good solutions in real-time environments. Lower-bound–based algorithms for the minimum tardiness-cost train-dispatching problem can be used to handle larger traffic volumes and cover longer planning horizons and larger dispatching territories than is possible with the current state of the art.

Interest in optimal train-dispatching systems has been revived in recent years by the development of Advanced Train Control Systems (ATCS) technology (*1–3*). Besides improved safety, the potential for lower fuel consumption and improved railroad operations are often cited as the major benefits of ATCS. However, there seems to be a lack of a conceptual framework defining the goals and flow of information between the optimal dispatching system and other components of ATCS that would lead to improved railroad operations.

TRAIN-DISPATCHING PROCESS

Train dispatching is a demanding and complex task. The survey paper by Petersen et al. (*4*) gives a brief description of the issues involved. Dispatchers monitor and control the movements of trains over railway lines and resolve potential conflicts between trains. The primary conflicts arising on single-track lines with passing sidings and partially double-track lines (such lines represent more than 90 percent of all railway lines in the United States) are meets of trains going in opposite directions. Two trains traveling in opposite directions cannot occupy the same single-track segment at the same time, or a collision would occur. Meets are resolved by switching one of the opposing trains onto a side track, where it waits until the other train passes. On fully double-track or multiple-track lines, there are no conflicts between trains going in opposite directions, but conflicts may arise between a fast train and a slow train that is traveling in the same direction in front of the faster train. Such conflicts can be resolved by overtaking, that is, by switching the faster train onto a parallel track on which it can pass the slower train, provided that this parallel track is not occupied by a train going in the opposite direction. Overtaking can also be done on a single-track line by switching the slow train onto a passing siding clear of the main track, where it stops and waits to be passed by the faster train. The number of ways in which conflicting train movements can be resolved is an exponential function of the number of trains and track segments.

Other duties of a train dispatcher include safe coordination of movements of roadway maintenance gangs, signal maintenance crews, and industrial switch engines, as well as a host of clerical duties concerned with maintenance of various operating statistics. Quoting from Sauder and Westerman (*5*), "Safety is the paramount consideration in all of the dispatcher's tasks."

Besides safety concerns, train dispatching is of paramount importance in the operation of a railroad network for another reason: dispatching decisions, through meet/pass delays, greatly influence train transit times and on-time performance. According to one study (*4*), 45 percent of the variance of train arrival times is due to the variance in over-the-line transit times. Unfortunately, dispatchers do not have at their disposal the information that shows systemwide effects of their decisions; their main incentives (besides safety) are to avoid delaying a "hot" (high-priority) train. As reported by Sauder and Westerman (*5*), a common response of dispatchers was to clear low-priority trains into a siding far in advance of incoming hot trains, thus minimizing the chance of delaying such a train while causing unnecessary delays for low-priority trains. During periods of very dense traffic, this strategy can often backfire. Delaying a cluster of low-priority trains can soon create an area of congestion in which all trains are delayed regardless of their priority.

Due to a heavy workload and insufficient information concerning future traffic, dispatchers are forced to cope with incoming traffic as it arrives and have little ability to make plans; that is, the function of a train dispatcher, is, at present, *reactive* rather than *proactive*. There is, however, a methodology designed to put dispatchers in a more active role in which they would work toward the common operating objectives of a railroad system, without increasing their workload.

D. Jovanović, Network Planning, Burlington Northern Railroad, 777 Main Street, Fort Worth, Tex. 76102-5304. P. T. Harker, Fishman-Davidson Center for the Study of the Service Sector, The Wharton School, University of Pennsylvania, Philadelphia, Pa. 19104-6366.

STATE OF THE ART IN OPTIMAL TRAIN DISPATCHING

In their article, Petersen et al. (*4*) gave a survey of computer-aided dispatching (CAD). The authors note that most of the information-gathering and record-keeping activities (on which dispatchers once spent 75 to 80 percent of their time) and other routine tasks can be computerized, and that commercial systems to handle these tasks are available. Petersen et al. (*4*) state that, as new CAD technology becomes available to simplify the dispatcher's tasks (such as automated signal clearing and calculations of expected train arrival times), the next important step is the development of optimal train-dispatching systems. The biggest obstacle towards development of such a system (*4*) is the combinatorial nature of the problem; for example, even for moderate traffic intensities and an 8-hr time horizon, it could take 12 days for a supercomputer to evaluate all possible meet/pass plans. Similar problems are reported in the only published account of a computer-aided optimal train-dispatching system implemented in the United States beyond the testing stage (*5*). Algorithms that significantly improve on the existing enumeration-based optimal dispatching algorithms were recently developed by Jovanović (*6*); the new algorithms allow for real-world dispatching problems to be solved optimally or near-optimally in less than a minute, thus removing the obstacle described above for most practical purposes.

However, an issue that has received little attention in the literature or from the vendors of CAD equipment concerns the goal of or the potential benefits from an optimal CAD system. As was lucidly noted in the paper by Duncan et al. (*7*), which describes the development and testing of an optimal CAD and train control system on an ore-hauling Australian railroad, "The difficulty in achieving the 'best' meet/pass plan is defining what is meant by 'best.' " Minimizing the sum of weighted train delays is the objective that is encountered most often; e.g., this objective was proposed in the pioneering work by Szpigel (*8*) and in the survey paper by Petersen et al. (*4*). The objective of the Norfolk Southern system is to minimize the sum of train delays and the priority-weighted sum of train lateness. A similar objective is proposed by Rockwell International (*9*) for the meet/pass planner under development within the Advanced Railroad Electronics System (ARES) project: in addition to minimizing the weighted sum of train delay and lateness, the ARES meet/pass planner attempts to minimize fuel consumption as well. The objective of the "driver-assist" optimal CAD system under development in Australia (*7*) is to minimize some combination of fuel consumption and train delay; however, the authors emphasize that railroad management must supply the objective or the "strategy" the system attempts to implement. Finally, the Union Switch and Signal CAD system, recently installed at CSX (*10*) and Union Pacific, does not have any explicit objective function in its meet/pass planner. Issues of how to calculate the weights or costs associated with train delay and lateness, whether they are always linear and constant in time and by train type, and what reference times should be used for the calculation of lateness are seldom discussed in the published literature.

COST MINIMIZED BY AN OPTIMAL COMPUTER-AIDED TRAIN-DISPATCHING SYSTEM

Fuel Consumption

A decrease in fuel consumption has often been mentioned as one of the major benefits of ATCS. Besides a CAD system that determines target times for trains as they traverse the line, it is necessary to have on-board train controllers that monitor train movements and advise the engineer on the most fuel-efficient manner for reaching the next target point within the allotted time. (We assume that the problem of fuel-efficient control of an individual train can be solved; see the report by Milroy et al. (*11*) for a promising approach.) Note that some fuel savings can be achieved through the use of on-board controllers, even if the minimization of fuel consumption is not incorporated into the CAD system objective function, as long as the planned meet/pass times are passed to the train controllers. A proprietary 1989 study based on research by the authors estimates these fuel savings at about 2.5 percent. To achieve greater fuel savings, it is necessary to trade off fuel versus train transit time and to incorporate fuel consumption minimization in the objective function of the CAD system. In the same proprietary study, it was established that the value of time lost by the rolling stock and lading of a train would be higher than the value of the fuel saved by decreasing the speed of the train in almost every instance. In another part of the study, a nonlinear optimization algorithm described by Kraay et al. (*12*) was applied to historical dispatching data sets with the objective of minimizing fuel consumption subject to the constraint that no train should arrive at its destination later than it actually did; under these conditions, the average fuel saving was about 7 percent. The main reason for the relatively low potential fuel savings lies in the shape of the fuel consumption curve and the ratio of train weights: those trains that incur the largest amount of dispatching delay are mainly the heavy and slow bulk-commodity trains whose fuel consumption is relatively insensitive to a decrease in speed, while most of the trains whose fuel consumption curve is sensitive to speed reduction are the fast high-priority trains that usually incur fewer dispatching delays.

An indirect benefit associated with decreased fuel consumption is decreased rolling stock and right-of-way wear and tear resulting from lower train speeds and the lower braking and tractive forces.

Another issue associated with the decrease in fuel consumption is that of the "robustness" of meet/pass plans. A train that is slowed down by the on-board controller to save fuel is more likely to be late for the planned meet than were it running at full throttle. This lateness may cause delay to the other meeting train, which was planned to go through the meet without any delay; the unanticipated delay can have a domino effect on the other meets and, thus, make the entire plan invalid. This trade-off between fuel and the reliability of a meet/pass plan, which directly influences the reliability of the trains' on-time performance, points to the need for the reliability of on-time arrival at the planned meet-point to be a primary goal of the on-board train controller; this issue has received very little attention in the literature.

It can be concluded that although at least 2.5 percent of total fuel consumption can be saved through the use of on-board train controllers coordinated by the CAD system, the reduction of fuel consumption should not be the primary goal of an optimal CAD system because of the high value of train transit time that must be traded for fuel. Once an optimal meet/pass plan is chosen, it may be possible to modify slightly the planned target times to achieve additional fuel savings as long as this does not significantly decrease the probability that the chosen meet/pass plan can be achieved. This approach can bring the total fuel saving up to 7 percent, depending on the desired trade-off between fuel cost and increased running times.

Cost of Train Delays

It should be noted that on a fully signalized railroad territory, there is no great need for new technology (such as ATCS), other than software, to achieve reductions in train tardiness and delay. Although the installation of on-board controllers and train-positioning systems envisioned by ATCS is necessary to realize fuel savings, adequate information and control capabilities may be provided initially by the existing signal system.

The ability of CAD systems to decrease significantly the amount of over-the-line delay that trains incur waiting for meets or overtakes or for a preceding train has been proven in practice and by numerical experiments. Since the optimal CAD system described by Sauder and Westerman (*5*) was installed by the Norfolk Southern Railroad, the delay per train has decreased by an average of 12.4 percent during the first year of operation and by 25.4 percent during the second year of operation when compared with the year preceding installation. The proprietary study of benefits from optimal CAD systems estimates (through the analysis of historical dispatching data) that train over-the-line transit times can be decreased by an average of 17 percent without involving any overtaking. The time savings that could be realized by individual trains vary greatly, depending on the amount of delay incurred and the importance (weight) assigned to a particular train in the CAD objective function; for example, the delay of several high-priority intermodal and mixed-freight trains was decreased by 1 to 3 hr out of a 9-hr transit time, with little additional delay to the other trains.

The goal of minimizing train delay and tardiness, suggested by most developers of optimal CAD systems, implies that certain costs are incurred by delaying a train and that certain benefits are gained if a train arrives early. We will attempt to analyze the potential sources of the implied costs and benefits.

Shipment Transit Time

The time when a train arrives at its next terminal can influence the total shipment transit time in two ways. First, if this terminal is the final destination for the shipment (e.g., in the case of passenger, intermodal, and bulk-commodity unit trains), then shipment tardiness is a direct function of the train arrival time. Second, if the car containing the shipment is scheduled to be transferred to another train, the late arrival of the inbound train relative to the departure time of the outbound train can increase the total shipment transit time by hours or days, depending on when the next appropriate outbound train will depart.

The actual cost of increased shipment transit time varies greatly with the type of shipment and can be very hard to quantify; the shipping contract may or may not require the railroad to pay penalties for late deliveries and the amount and form (linear, progressive) of this payment varies. Another, less tangible cost associated with late or unreliable shipment deliveries is customer dissatisfaction and loss of business to other carriers or, inversely, the loss of potential revenue that the customer would be willing to pay if the quality of service was higher. These costs depend on the customer's sensitivity to disruptions to his distribution and/or production process (which can be quite high if he is using just-in-time inventory control), as well as on the time-value of the commodity being shipped.

In most cases, there are no benefits from delivering the shipment early; on the contrary, the customer may not be ready to receive it and early delivery can cause unnecessary congestion in either the customer's receiving and inventory system or in the railroad terminal. The latter case can be observed when trains arrive before the planned time at a classification yard, occupying capacity planned for other trains and causing longer processing times and late departure of those trains.

The above discussion shows that the cost of train delay with regard to shipment transit time can only be calculated relative to some *scheduled* time: either the scheduled shipment delivery time (the train's published scheduled arrival time in the case of passenger and intermodal trains) or the latest arrival time necessary for the shipment to make its connection. The cost of shipment transit time for a train composed of several blocks of cars, each block destined for a different outbound train, will most likely be a nondecreasing staircase-shaped function of train arrival time, with each step corresponding to the connection time of a group of shipments (i.e., a block of cars) destined for the same outbound train.

Rolling Stock Value

It is obvious that significant reductions in rolling-stock capital cost could be achieved if the turn-around times of trains were decreased, because a smaller number of cars and locomotives could produce the same output, or the same equipment could produce a higher output. These reductions, however, can only be realized if the *planned* or *scheduled* train transit times are decreased to take advantage of the faster train movements made possible by optimal CAD systems. For those trains that are not scheduled (e.g., unit trains), the target arrival and departure times should be set in real time by a systemwide operating plan and passed down to dispatchers as objectives, alongside target arrival/departure times for scheduled trains. Only some local trains may be left to the complete discretion of the dispatchers with regard to their arrival and departure times. Thus, the goal of increased fleet utilization cannot be directly incorporated within the CAD objective function at the real-time level, because it belongs within the systemwide train scheduling process. Equipment-related costs that could be incorporated within the CAD objective function are related to the exceptions from the schedule. (For example, the inability to form or adequately power a new train and/or the lack of empty cars to be delivered to the customer may be caused by the late arrival of trains whose rolling stock was scheduled to be used for this purpose.) These costs are related to the planned train arrival time, which can be changed in the operational plan to be earlier than the published scheduled time if the train is bringing equipment (cars, locomotives) urgently needed for another train. On the other hand, *unplanned* early arrival does not bring any benefit in terms of equipment utilization and may cause yard congestion.

Crew Costs

The wages paid to train crews for a particular leg of a train trip depend on the type of agreement between the labor union and the railroad; this discussion deals only with those crew-related costs that are a function of the train's arrival time. Although the current practice in North America is to pay crews primarily be the mileage of the trip, it is not inconceivable that they could be paid for the actual hours worked or for the scheduled train travel time plus overtime pay if the train is late. Thus, crew-related costs are likely to increase with late train arrivals, either through direct pay or through worsened labor relations.

Another important crew-related cost is associated with the federal rule that prohibits crews from operating a train after 12 continuous hours spent on duty; when excessive train delays over the scheduled travel time cause this limit to be reached, the train must be stopped on the line and a fresh crew brought in. The cost of bringing in the relief crew and the resulting delays to the train and the blockage of the line should be incorporated within a CAD system.

OBJECTIVE OF AN OPTIMAL CAD SYSTEM

In the previous section it was argued that, at the operational level, the costs related to train transit time could be defined only with reference to some scheduled or planned target arrival times. These target times are not necessarily the published schedule times, in part because not all freight trains have published scheduled arrival times, and in part because various disturbances in the schedule may require a new operating plan that differs from the tactical (published) schedules. For example, if a high-priority intermodal train were delayed during previous legs of its itinerary to the point where it could not reach its final destination on time, then its planned target arrival time must be shifted forward, and the plan for other affected trains (e.g., those waiting for the locomotives from the late train) must be adjusted accordingly.

In the case of unscheduled trains (e.g., unit coal trains), once a decision is made to run such a train, the desired running time and the desired arrival time at its destination should be made a part of the plan, and shipment-, equipment-, and crew-related costs of exceeding the planned time could be known. For every train planned to enter the dispatcher's territory the operational cost as a function of the train's arrival time could be known.

The objective of an optimal CAD system is to ensure the implementation of the systemwide operating plan over a given dispatching territory or, if necessary, to ensure that the fewest cost deviations from this plan are achieved. Thus, rather than being governed by standard operating procedures that state, for example, that a low-priority train should always be put on a siding when meeting a high-priority train, dispatchers will be guided by the CAD-generated meet/pass plans that minimize systemwide costs even if that means delaying a high-priority train running early in order to get a medium-priority mixed-freight train to arrive on time for a connection. The importance of high-priority trains is maintained through the high costs associated with the *late* arrival of such trains rather than by fixed operating rules. It should be noted that track maintenance work can also be assigned a cost function and be scheduled like a train, admittedly a very slow one. Rather than putting the track to be maintained out of service at the convenience of the maintenance-of-way gangs, dispatchers could evaluate possible "windows" of the required length and choose those that disturb the traffic flow the least.

A systemwide operating plan, then, consists of target arrival times for all trains in the system, planned track maintenance, planned car-block transfers at the yards, planned distribution of power and empty cars, crew rosters, etc., which should be updated periodically using feedback on the train status as generated by the CAD and yard information systems. Such operating plans are used by railroads at present, except that instead of setting the target train arrival times for the dispatchers, the operating plan is based on the expected time of arrival. The proposed methodology aims to change the position of railroad management, dispatchers, and other field officers; rather than being reactive, they can be proactive by setting clear operating objectives in the form of a systemwide operating plan and by minimizing the systemwide effect of disturbances when these objectives cannot be realized.

Note that this methodology does not require all trains to be scheduled at the tactical level. The only requirement of this methodology concerning the operating plan is that it should be realistic (i.e., feasible). The more "robust" the tactical schedules are, the easier it is to maintain a feasible operating plan and vice versa; if the tactical schedules are infeasible, then an operating plan based on those schedules would be impossible to maintain. Note that the feasibility of an operating plan should not be evaluated solely on the basis of over-the-line operations, but also by considering yard operations and capacity, systemwide flow of locomotives and empty cars, etc.

One of the main benefits of the use of optimal CAD systems within this framework will be the increased reliability of service through (a) better on-time performance for high-priority trains, (b) fewer missed connections and less deviation in the transit time for mixed-freight shipments, (c) faster turn-around and more predictable service of bulk-commodity unit trains, and (d) more evenly distributed workload and more reliable operations at yards. An example from the proprietary study of benefits of optimal CAD systems can be used to illustrate the potential of well-designed CAD systems to improve on-time performance. During 16 hr of operations on a 300-mi-long line, 7 high- and medium-priority trains out of 25 total trains accumulated a total of 724 min of tardiness, ranging from 3 hr to 40 min per train; if these trains were dispatched with the support of an optimal CAD system, the total amount of tardiness for all 25 trains could have been only 14 min.

DESIRED FEATURES OF AN OPTIMAL CAD SYSTEM

Flexibility and Speed of the Optimal Meet/Pass Planning Algorithm

The optimal dispatching algorithm should be able to handle various types of cost function associated with train lateness; some obvious examples include piecewise linear, piecewise quadratic, and step functions, none of which are differentiable. The lower-bound–based algorithms described by Jovanović (*6*) can handle any nondecreasing function of train arrival times. (The constraint on the monotonicity of the cost functions is not overly restrictive: trains simply can be constrained not to arrive before a certain time, subject to congestion levels in the terminals.) In addition, the meet/pass planning algorithm should be able to dispatch trains between any two points on the line, even if there are no terminals at these points, in order to be able to handle maintenance-of-way gangs as special trains.

Dispatchers should also be allowed to exercise their judgment and expertise by manually constraining certain train meets to particular points or by preventing some meets from occurring at certain points. Finally, the algorithm should be fast enough to provide reasonably good solutions in enough time to allow the dispatchers to test the sensitivity of the plan to some random events, and to generate a revised meet/pass plan quickly if some of the input parameters change (e.g., if one of the locomotives in a certain train has broken down and this train can no longer achieve planned speed). At the same time, to allow systemwide operational planning, the algorithm should be able to handle long time horizons (at least 12 hr) and large dispatching territories. For example, if trains take on the average 9 to 10 hr to transit a particular line, a meet/pass planning horizon of 8 hr is obviously inadequate, because it cannot guarantee on-time arrival of all trains that are now in the system, let alone be used to plan for trains that have not yet entered the territory.

The heuristic algorithm described by Jovanović (*6*) is shown to satisfy all of the above requirements.

Display of Information

Time-distance diagrams ("string-line" diagrams, in railroad jargon) have traditionally been used by railroads to depict the progress of trains over the line. Thus, it is natural, as in most proposed CAD systems, to use time-distance diagrams to communicate suggested meet/pass plans to the dispatchers. One such diagram depicting a meet/pass plan over a predominantly single-track Whitefish to Spokane line is shown in Figure 1; the current plan time is 4:00 a.m. and the diagram shows planned train movements until midnight—the end of planning horizon. A track schematic is given along the left side of the diagram with sidings and double-track sections represented by rectangles. (There are only two short double-track sections in this example, one from Sandpoint to Algoma and the other one from Irvin to Spokane.) All train meets occur either at a siding or over a double-track section.

The information content presented to the dispatcher in the meet/pass diagram is very important. To illustrate this point, compare Figures 1 and 2. At first glance, the two plans are almost identical; yet the total amount of train lateness for the first plan is 230 min versus 40 min for the second plan. The cost of lateness (e.g., missed connections) is more than three times higher for the plan in Figure 1 than for the plan in Figure 2 (the latter plan was optimized using algorithms described by Jovanović (*6*)). All this is not obvious from the diagram: a relatively small change in meet locations for some train pairs, such as those marked with circles in Figure 1, eliminated or reduced late arrivals for six trains. The difference becomes more apparent if one observes that late trains in both diagrams are drawn using dashed lines from the point where they were made late; the number of late trains has decreased from seven in Figure 1 to only two in Figure 2. It would be even easier to differentiate the meet/pass plans if the trains were drawn with increasingly intense shades of red corresponding to the increasing cost of the trains' lateness.

The train-tardiness cost information embedded in the graphical interface of an optimal CAD system clearly presents to the dispatchers the effects of their actions and how well they are realizing the objectives set for them. In this way, the dispatchers can work toward achieving the common systemwide operating plan rather than attempting to move trains from their territory as fast as they can just so the trains can become somebody else's responsibility.

Assigning Track Time to Maintenance-of-Way Work Gangs

Maintenance-of-way (MOW) gangs could be treated as special trains within the context of an optimal CAD system because they occupy track capacity in space and time and cause congestion in a similar manner as trains. The main scheduling difference between trains and MOW gangs is that the activities of the latter are much less time sensitive; i.e., the value of the output of MOW gangs often remains constant when the completion time of the activity is shifted a few hours later or earlier; the same is not true for trains. The only direct costs associated with track maintenance that are time sensitive are those connected to labor (including nighttime and overtime) and equipment utilization. These costs are often not high enough to justify the hours of lateness caused to high-priority trains, but they may be high enough to give higher priority to the MOW gang over a low-priority coal-hauling train. Train dispatchers should have the final say regarding the exact time windows assigned to the MOW gangs and should be able to assess the effects that the assignments will have on train performance.

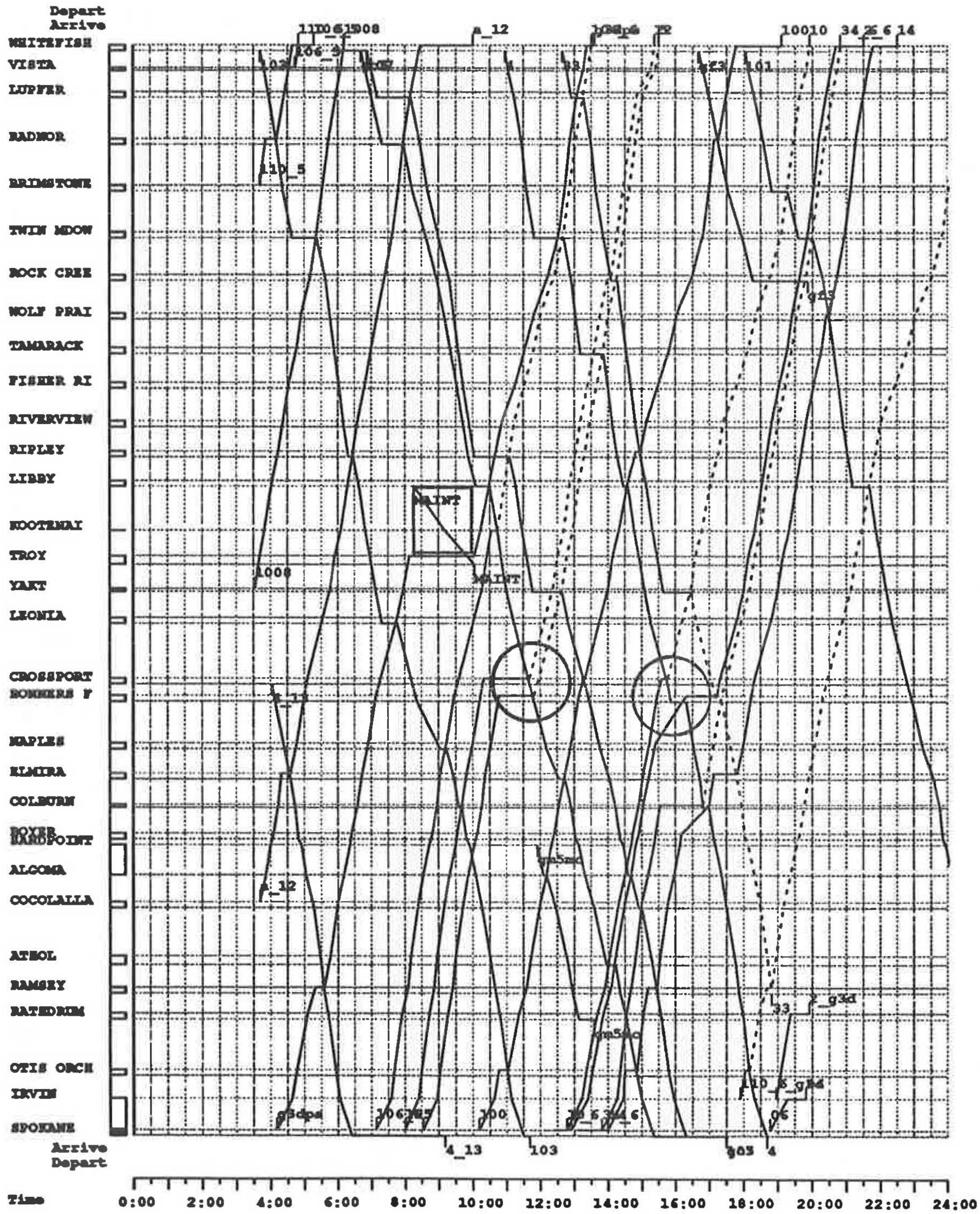

FIGURE 1 Unoptimized meet-pass plan.

The importance of the efficient allocation of time for track maintenance and its optimization by a CAD system is illustrated in Figures 2 and 3. In Figure 2, the outage of the track segment between Libby and Troy, represented by a rectangle marked "MAINT" that occupies this segment between 8:10 and 10:00 a.m., produces a significant delay and 30 min of lateness to Train "g3dps" (this is the train that departs from Spokane toward Whitefish starting shortly after 4:00 a.m.). If the maintenance block were treated as a special train that could be shifted in time rather than as a given constraint and the interval allocated to maintenance were shifted just 15 min ahead to 8:25 a.m. (10:15 as shown in Figure 3), Train "g3dps" could traverse the Libby–Troy segment before it was closed for maintenance and arrive on time at Whitefish.

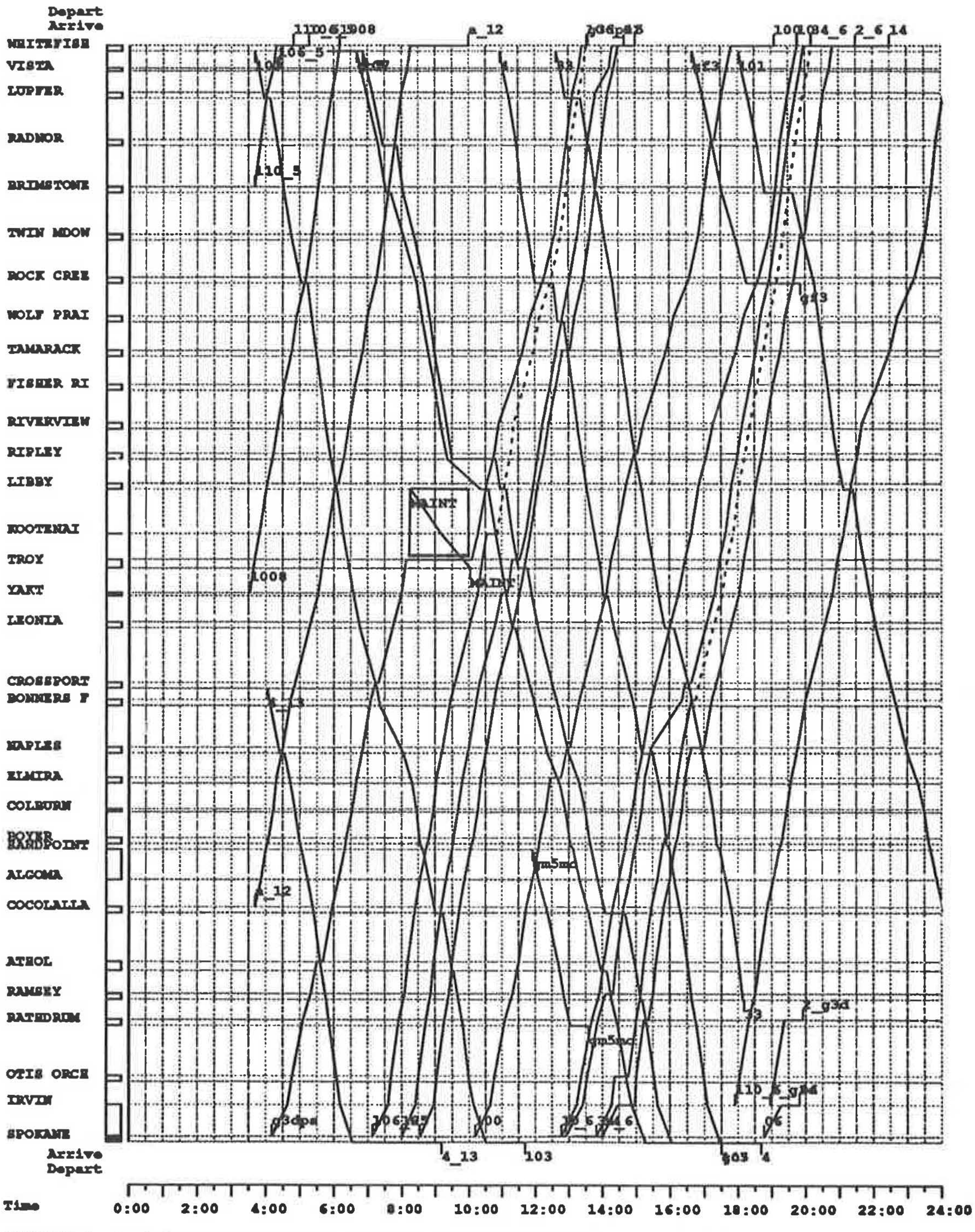

FIGURE 2 Optimized meet-pass plan with track maintenance treated as a constraint.

Control over Trains Entering the Line

In addition to having the control over time and track segments assigned to MOW gangs, dispatchers should control when and in what order new trains are permitted to enter their territories. The present situation in which dispatchers very often do not know when and which trains are going to enter their territory at a junction or from another line is analogous to air-traffic controllers having new aircraft appear unexpectedly in the middle of the air space that they control.

A similar argument can be applied to the order in which trains entering the line leave the yard; letting a slow train with a loose schedule depart in front of a fast high-priority train with a tight schedule either will delay the faster train and probably make it arrive late or require the faster train to overtake the slower train in a time-consuming and track

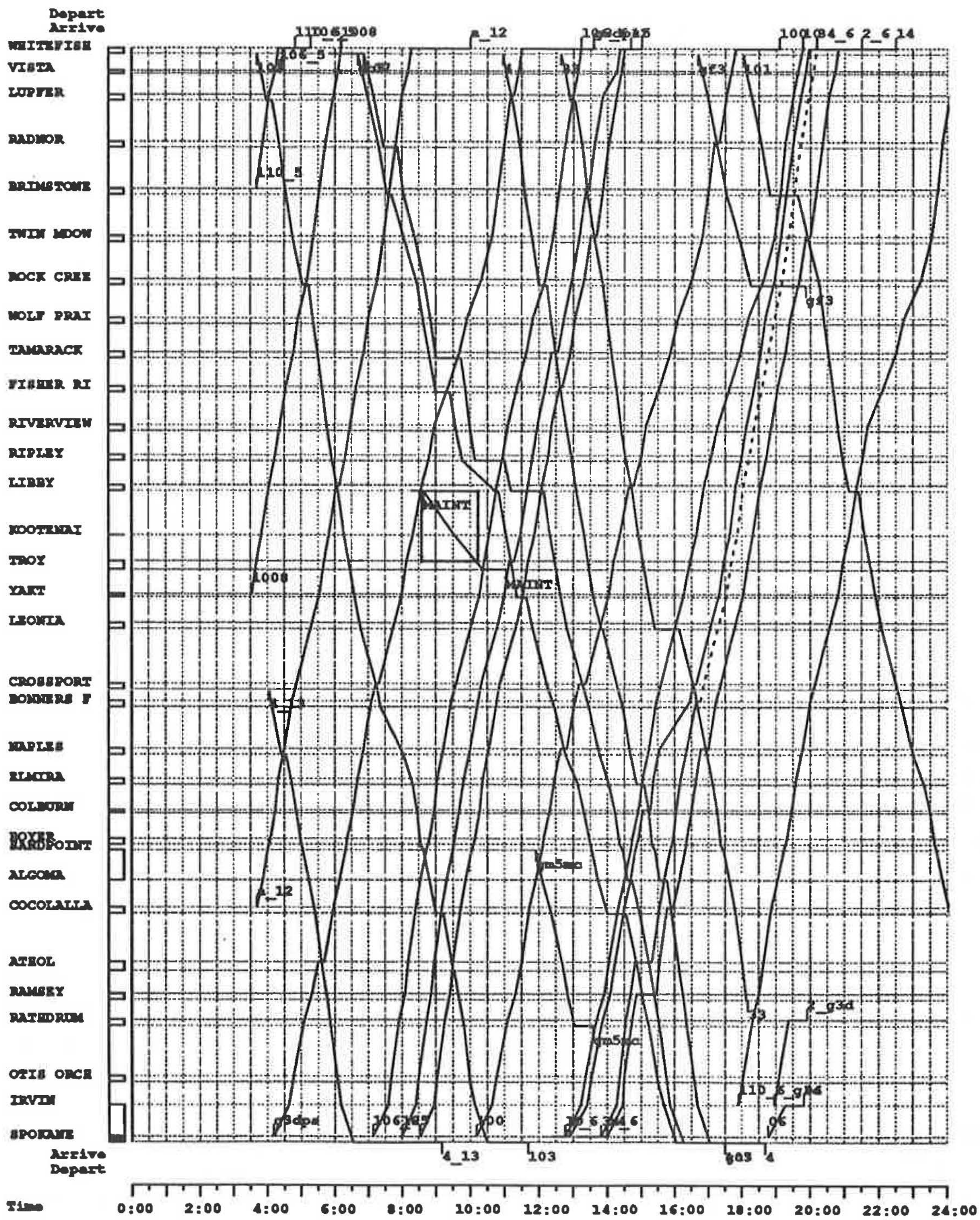

FIGURE 3 Optimized meet-pass plan with track maintenance scheduled as a train.

capacity-intensive process. However, delaying a train that is ready to depart from a yard so that some other trains can depart from the yard first can cause congestion in the yard. Hence a trade-off between line delays and yard delays may be required.

The importance of the ordering of trains departing from the yard is shown in Figures 3 and 4. The only late train in the meet/pass plan in Figure 3 (dashed line) is delayed and made late by a slower train in front of it. Although the late train departs from Spokane almost 1 hr after the preceding slow train (1:00 p.m. versus 1:45 p.m.), this time gap is quickly eliminated by a difference in speed between the two trains. Normal railroad practice would be to allow the late train to overtake the slow train in front; however, overtakes can often be eliminated by planning the order in which trains depart from terminals. In Figure 4, the slow train in front was ordered to depart from Spokane at 2:55 p.m., behind the train that is late in Figure 3, and both trains arrive on time at Whitefish.

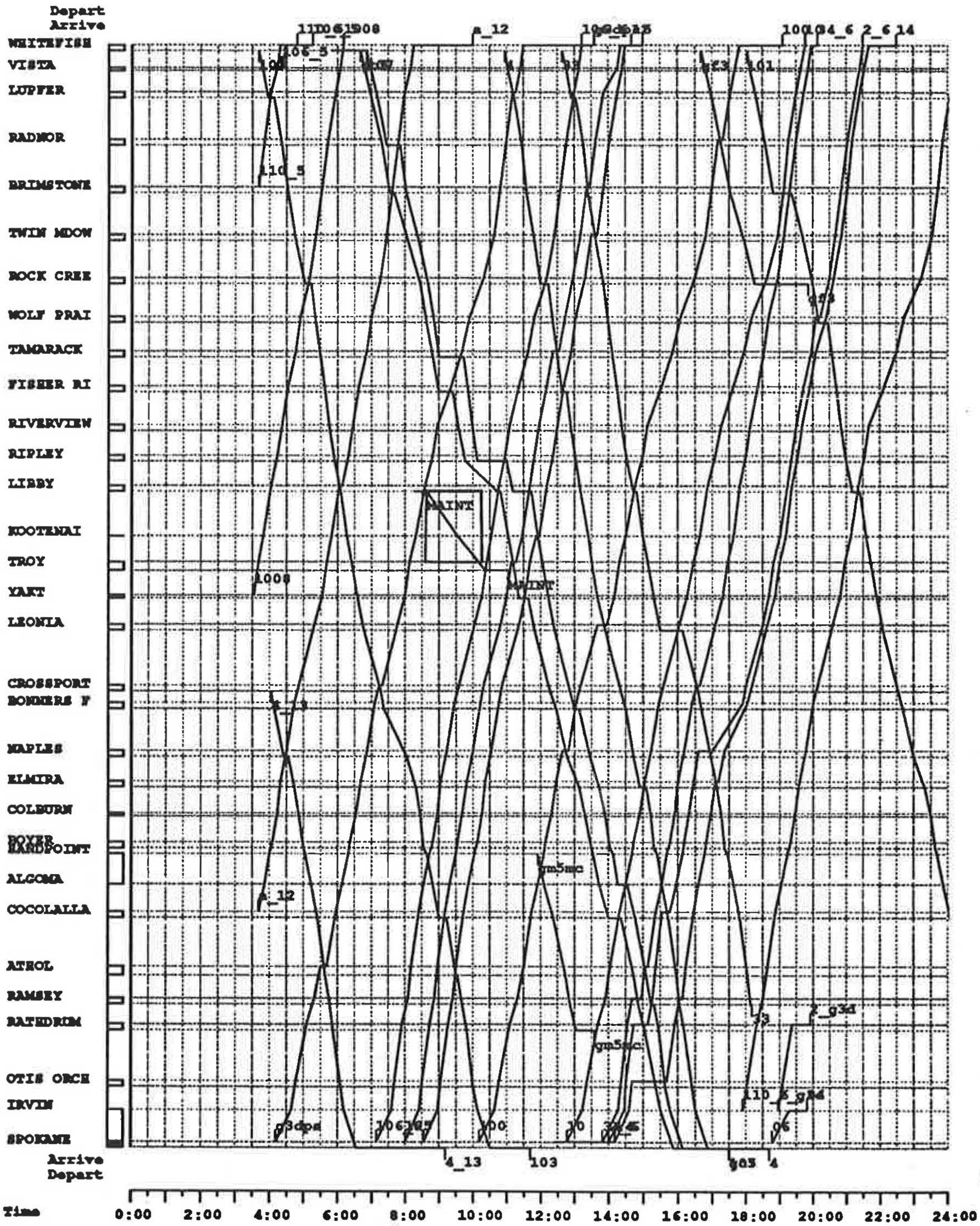

FIGURE 4 Eliminating late arrivals by reversing order of departure.

If this change in departure time is planned far enough in advance and there is enough yard capacity, almost 1 hr of the slow train's on-duty crew time can be saved along with gaining more time to assemble, power, and inspect the train.

One possible approach toward incorporating the line/yard trade-offs would be to include a simple yard model in the CAD algorithm. This model would account for the limited yard storage capacity by imposing additional constraints and treat the line, branches, and yards as a continuous dispatching territory covered by a single optimal CAD system. The algorithms described by Jovanović (*6*) can be extended to accommodate this approach. Another, probably less optimal approach would be to optimize yard and branch line plans separately and then have some higher-level model (or a decision-maker) modify these plans to ensure their compatability.

EXTENSIONS

The CAD methodology described in the previous sections assumes that no random events occur during the planning horizon. Of course, this is not realistic; equipment failures and adverse weather conditions influence the performance of trains and track availability, which in turn can render the current meet/pass plan infeasible. Another source of randomness in line-haul operations is the late departure of trains from their originating yards or the addition of new trains. Thus, the further one extends the planning horizon of a CAD system, the greater the probability is that the plan will have to be modified due to unforeseen events. There are several ways to approach this problem.

The approach used in the Norfolk Southern optimal CAD system (*5*) is to discount future costs associated with dispatching delays; i.e., each train's cost is multiplied by a discount factor $(T - a)/T$, where T is the planning horizon of the meet/pass planner and a is the interval between the current time and the expected time of the train's arrival in the territory ($a = 0$ for trains already on the line). One problem with this approach is that it tends to underestimate costs; i.e., although it is true that the actual meet/pass pattern to be implemented in the future and the associated costs are uncertain, it is much more likely that the actual costs and delays will be *higher* than the costs predicted by the optimal meet/pass plan because of potential accidents and the deterioration of the dynamic performance of some trains. Nondiscounted cost estimates associated with a minimum-cost meet/pass plan will already tend to be below the expected cost; the use of discounted costs will amplify this error.

Another approach might be to focus on the nature and causes of the stochastic events that affect line-haul operations. A belief has started to emerge in the railroad industry that a large number of these events can be controlled and significantly reduced or eliminated by better planning, preventive maintenance, and better work discipline. The examples of high-precision, punctual operations of Japanese, French, and Swiss railroads show that random events can be controlled and accounted for. The highly reliable operations of these railroads can be attributed in part to fully double- or multiple-track lines with fewer train conflicts; in fact, a well-designed CAD system can have an effect similar to that of adding an additional track in terms of the reduction of train conflicts. However, some practical way to handle the problem of stochastic events is needed before the long-term efforts aimed at better planning, maintenance, etc., produce the desired results. The reliability of a meet/pass plan can be increased by allowing sufficient slack in the plan. Thus, instead of decreasing the *average* train transit times, the capabilities of newly installed optimal CAD systems can be used to decrease the *variance* of train transit time. With sufficient reserves built into the minimal point-to-point train running times used in the meet/pass algorithm, the probability that trains will be able to achieve these times increases. As the random events become more controllable, the slack in the meet/pass plan can be reduced and, consequently, the scheduled transit times can be decreased to yield faster service and better rolling stock utilization.

Another important issue related to the implementation of optimal CAD systems is the problem of coordination of dispatching and yard activities throughout the railroad network through the generation of targets and objectives for the dispatchers and yardmasters. A systemwide, real-time information system containing positional data and status of all carblocks and trains is necessary to achieve the full benefit from optimal CAD systems. A need and a research opportunity arise in providing additional modeling and optimization tools to support the generation of systemwide operating plans. Only through efficient systemwide operational planning of railroad operations can the potential of the proposed ATCS technology be fully realized.

ACKNOWLEDGMENTS

This research was supported by a grant to the University of Pennsylvania by the Burlington Northern Railroad and by the National Science Foundation under their Presidential Young Investigator Program.

REFERENCES

1. *High Technology Rides the Rails*. Association of American Railroads, Washington, D.C., 1985.
2. *Revolution in Research Will Benefit Railroads and Shippers*. Association of American Railroads, Washington, D.C., 1986.
3. G. Welty. ATCS: On Time, on Target. *Railway Age*, June 1987.
4. E. R. Petersen, A. J. Taylor, and C. D. Martland. An Introduction to Computer-Assisted Train Dispatching. *Journal of Advanced Transportation*, Vol. 20, 1986, pp. 63–72.
5. R. L. Sauder and W. M. Westerman. Computer-Aided Train Dispatching: Decision Support Through Optimization. *Interfaces*, Vol. 13, 1983, pp. 24–37.
6. D. Jovanović. *Improving Railroad On-Time Performance: Models, Algorithms and Applications*. Ph.D. dissertation. Department of Systems, University of Pennsylvania, Philadelphia, 1989.
7. I. B. Duncan, K. M. Winch, and G. A. Bundell. Driver-Assist—Microprocessor Technology to Aid in the Scheduling of Trains. In *Computers in Railway Management: Proceedings of COMPRAIL 87 Conference in Frankfurt, West Germany* (T. K. S. Murthy, L. S. Lawrence and R. E. Rivier, eds.), Springer-Verlag, Berlin, Germany, 1987.
8. B. Szpigel. Optimal Train Scheduling on a Single Track Railway. In *Operational Research '72* (M. Ross, ed.) North-Holland, Amsterdam, Netherlands, 1973.
9. R. D. Burns and D. B. Turner. Safety and Productivity Improvement of Railroad Operations by Advanced Train Control Systems. In *Proceedings of a 1989 Joint IEEE/ASME Railroad Conference*, Philadelphia, Pa., April 1989.
10. CSX: 'Welcome to the 21st Century!' *Railway Age*, June 1989.
11. I. P. Milroy, B. R. Benjamin, R. L. Payne, J. Kautsky, P. J. Pudney, A. J. Skinner, and S. P. Rice. *Freight Locomotive Fuel Conservation Project—Final Report*. South Australian Institute of Technology, The Levels, Australia, 1987.
12. D. Kraay, P. T. Harker, and B. Chen. Optimal Pacing Of Trains in Freight Railroads: Model Formulation and Solution. *Operations Research*, 1990 (forthcoming).