An Estimate of the Punctuality Benefits of Automatic Operational Train Sequencing

Rien Gouweloos[1] and Maarten Bartholomeus[2]

[1] Atos Consulting, Papendorpseweg 93
3528BJ, Utrecht, The Netherlands
rien.gouweloos@atosorigin.com
http://www.atosorigin.com

[2] Holland Railconsult, Postbus 2855
3500GW, Utrecht, The Netherlands
mgpbartholomeus@hr.nl
http://www.hrn.nl

Abstract. In Dutch railway operations, most of the rescheduling decisions in the operational phase, following some disturbance, involve the resequencing of trains. These decisions are being taken using only approximate train information and operational rules. Improvements have been formulated but, since no insight in the potential gain in punctuality exists, lack a convincing business case. In this paper, using only elementary methods, we derive an estimate for this punctuality gain.

1 Introduction

Increasing the reliability of the Dutch Railways ranks high on the national political agenda and is therefore a high priority in the Railway sector. In fact the sectors strategy statement, called "Benutten en Bouwen" ("Utilize and Build"), makes this into its central tenet. This is a marked change from previous strategies, which concentrate on maximizing the volume of railway traffic per unit of infrastructure. Benutten en Bouwen takes a clear position: maximizing the utilisation of the infrastructure will be an illusionary goal, unless first reliability is improved drastically.

Given the many daily departures from the predefined timetable, it will be clear that the operational processes of rescheduling and dispatching train traffic are very busy indeed. Each day, thousands of minor and major rescheduling decisions are being taken. Obviously, the quality of these processes is of vital importance. Quantitatively however, very little is known about the influence of scheduling quality on reliability. Specifically, the question which part of the occurring unreliability is due to suboptimal rescheduling is unanswered. This paper addresses this question.

The dispatching process is far from being perfectly accurate. Train information, such as position (delay) and speed is only approximately available. Inaccuracy is further increased by the fact that, although traffic control in the Dutch

F. Geraets et al. (Eds.): Railway Optimization 2004, LNCS 4359, pp. 295–305, 2007.

railways heavily uses information systems, train dispatching is still exclusively a human decision. A conscious decision has been made to limit the role of IT systems to presenting timetable and train status information to human decision makers. Efforts to assist these people by more advanced IT scheduling solutions are yet to leave the prototype stage. Although there are some valid reasons to leave dispatching in human hands, we have to realize that, compared to computers, human beings are not very good at making rapid, consistent and precise rule-based decisions. Thus, errors due to inaccurate train information and human operator errors combine into an imperfect decision making process. In this paper, we will confront this situation with a situation with perfect decision-making, referred to as "automatic sequencing".

In the Netherlands, the operational processes of rescheduling train traffic have been structured into a number of control layers. We will deal with the lowest layer, which we will call the dispatching function. In this paper we view dispatching as a series of decisions on the order of trains on given routes: sequencing. In practice, this covers the vast majority of all relevant dispatching decisions.

Detailed instructions have been prepared for the dispatchers when (not) to change the order of trains from the order as given by the original timetable. All situations with one delayed train have been exhaustively tabulated nationwide. Essentially, these so-called if-then sheets contain the following statements:

If *train x has a delay between d_1 and d_2,* ***then*** *it should be given access to the infrastructure between trains y and z.*

Our calculation builds on these if-then sheets. Through the if-then sheets, contact with actual data is being made. They are the reason we can get a result without the need for explicit modelling of infrastructure or timetable information.

The operation uses if-then sheet information for dispatching. Obviously, an operator needs information on delays in order to make the correct decisions. Unfortunately, this information is not perfectly available. We estimate that delays used in the decision process are distributed around the real delay values with a standard deviation of 2 minutes. The reliability of railway traffic is commonly operationalised in terms of the punctuality of passenger traffic. In the Netherlands, punctuality is defined as the percentage of trains with a delay of three minutes or less at major stations, as compared to the original timetable. We will derive a ballpark estimate for the amount of dispunctuality incurred from current operational procedures and systems.

This paper is organized as follows. In Section 2 we present out formalism. In Section 3 we discuss the various inputs to our calculation, such as if-then sheets, delay distributions and decision making errors. Section 4 presents the results. In Section 5 we present our conclusions and offer some suggestions for follow up.

This work has been done as part of a project under the jurisdiction of Railverkeers-leiding. The project called for the preparation of a business case for the development of improved control systems for the dispatching layer. Railverkeersleiding, part of the public domain ProRail organization, is responsible for operational capacity management of the Dutch Railway Infrastructure. Railverkeersleiding performs rescheduling of the infrastructure under continuous consultation with the transport companies.

2 Approach

2.1 Elements of the Formalism

Loss Function

Let us consider two trains A and B approaching an insertion point P. The optimal order must be decided upon. In the following, the optimal order will be the order in which the summed delays of A and B at some reference point will be minimal. This reference point is usually a point somewhere after P, where the two trains start having separate routes. If A and B have the same characteristics (speed) after P, any point after P will do and our optimal order rule reduces to a first come first serve rule.

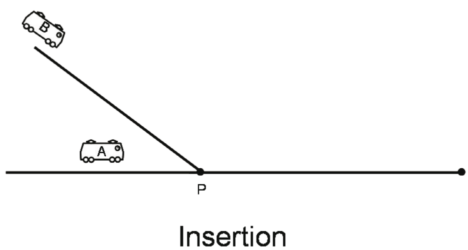

Fig. 1. An insertion point

We simplify the dynamics of the problem by assuming that if train A (B) is given priority at the insertion point while hindering train B (A), train B (A) always leaves the insertion point at a fixed headway H after A (B). In this case, there always exists one relative delay of A and B d_c for which the optimal order changes from AB to BA. We call this the characteristic delay. The if-then sheets tabulate these characteristic delays for all train pairs at insertion points in the Netherlands.

We call the amount of extra summed delay for A and B at the reference point resulting from the choice for a suboptimal order at P, the loss L. We can plot L as a function of $v = d_A - d_B - d_c$: L is a piecewise linear function of v, symmetric in $v = 0$. In the diagram we also show our definition of four "regimes" 0 to III in the values of v.

It is of interest to note that an optimal decision in the summed delay is not necessarily optimal for the contributing trains taken separately. Therefore, in individual cases it is possible that the optimal decision decreases punctuality.

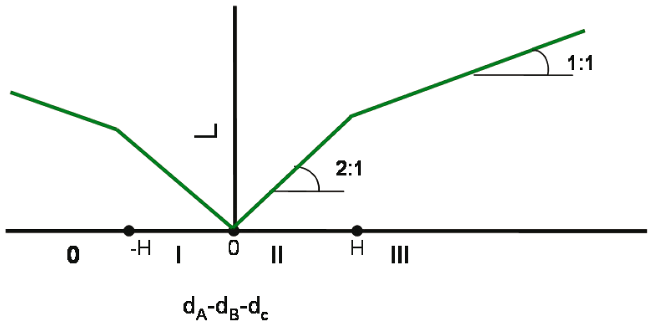

Fig. 2. Loss function

Delay Distribution
Trains approach the insertion point with a delay according to a density function $D(d)$. For this we use a conventional negative exponential fitted to the observed punctuality.

$$D(d) = 1/d_{\text{av}} \exp(-d/d_{\text{av}}) \quad (d > 0, 0 \text{ elsewhere}) \tag{1}$$

Note that in this we assume the same distribution will hold always and everywhere.

Control Error Distribution
Our purpose is to determine the punctuality effect of operator errors due to incorrect delay information. Let $F(d, d_g)$ be the probability density that a delay d_g will be assumed by the operator where actually a value d exists. For F we use a normal distribution

$$F(d, d_g) = F(d - d_g) = (1/(\sigma\sqrt{2\pi})) \exp(-(d - d_g)^2/2\sigma^2) \ . \tag{2}$$

Human operators tend to act conservatively, that is they tend to leave a planned train sequence in place, until it is very obvious that it should be changed. In terms of F, this behaviour means that F is no longer symmetric around $d - d_g$. In order to investigate these effects whilst leaving our formalism intact, we have introduced a parameter called the shift s:

$$F_s(d, d_g) = F_s(d - s - d_g) = (1/(\sigma\sqrt{2\pi})) \exp(-(d - s - d_g)^2/2\sigma^2) \ . \tag{3}$$

An erroneous decision will be made if a delay $d < d_c$ is assumed, while in fact $d_g > d_c$ and vice versa. For regime 0 (compare Figure 1) this chance is given by:

$$P_s(d - d_c) = \int_{d_c - d}^{\infty} F_s(d, d_g) dd_g = 1/2 \text{erfc}((d_c - d - s)/(\sigma\sqrt{2})) \ , \tag{4}$$

in which erfc is the complementary error function. Similar relations can be written for the case $d > d_c$ and $d_g < d_c$ and the other regimes.

2.2 Overall Punctuality Effect

We are interested in computing overall punctuality effects of suboptimal sequencing. Consider trains A and B with delays d_A and d_B respectively. Train A turns from being punctual to dispunctual when $d_A < 3$ while $L_A > 3 - d_A$. The reverse effect occurs if $d_A > 3$ while $L_A < 3 - d_A$.

Lets define the following probability distributions, all defined with respect to a sequencing point with characteristic delay d_c:

$P_1(d)$	The probability density for a train with delay d to be punctual,
$P_2(d_1, d_2)$	The probability density for trains with delays d_1 and d_2 to be incorrectly sequenced,
$P_3(d_1, d_2)$	The probability density that the extra delay for train 1 will exceed $3 - d_1$.

In order to keep this paragraph concise, we will only give explicit formulas for train A and regime 0 of the loss function. In this case:

$$P_1(d_A) = \Theta(3 - d_A) \tag{5}$$

$$P_2(d_A, d_B) = \text{erfc}(|(d_A - d_B) - d_c|/\sigma) \tag{6}$$

$$P_3(d_A, d_B) = \Theta(d_c + H - 3 + d_B) \tag{7}$$

where Θ is the Heaviside Step Function.

The probability train A turns from being punctual to dispunctual at a sequencing point with characteristic delay d_c is given by:

$$\int_{-\infty}^{\infty} dd_B V(d_B) \int_{-\infty}^{\infty} dd_A V(d_A) P_1(d_A) P_2(d_A, d_B) P_3(d_A, d_B) = \tag{8}$$

$$\int_{3-d_c-H}^{\infty} dd_B V(d_B) \int_{0}^{\min(3, d_c+d_B)} dd_A V(d_A) P(|d_A - d_B|) \ . \tag{9}$$

Similar expressions can be derived for

- the other regimes of the loss function;
- the probability that the train turns from being dispunctual to punctual;
- train B.

That is, all in all $4 * 2 * 2 = 16$ similarly structured double integrals have to be evaluated.

Lets call the grand total of these terms, the probability that an incorrect sequencing decision with characteristic time d_c turns some train into being dispunctual, $D(d_c)$.

It is straightforward to compute the chance of an incorrect decision E and the average value of the loss function L_{av}, for a decision with characteristic time d_c by:

$$E(d_c) = \int_{-\infty}^{\infty} dd_A \int_{-\infty}^{\infty} dd_B V(d_A) V(d_B) P(d_A - d_B - d_c) \tag{10}$$

$$L_{\text{av}}(d_c) = \int_{-\infty}^{\infty} dd_A \int_{-\infty}^{\infty} dd_B V(d_A) V(d_B) P(d_A - d_B - d_c) L(d_A - d_B - d_c). \tag{11}$$

In all cases (9), (10) and (11) an overall effect is found by summing over the number of potential daily resequencing decisions $A(d_c)$ as given by the if-then sheets. After suitable normalizations for the total number of trains and punctuality measurement points, overall effects result.

2.3 Correction Terms

Some corrections to the formalism outlined above have been considered. These corrections are not expected to be very accurate, but serve to get some feeling for the reliability of the results.

Delay may be Nullified Before Measurement
Trains are planned with a driving time margin. That is, if the train has a delay d somewhere between punctuality measurement points (nodes), it will be able to reduce this delay somewhat before measurement. From the timetable, we have determined an average value dn for the amount a train can reduce its delay before the next node. We use separate values for resequencing points just after leaving a station (driving time margin for the distance between stations) and resequencing points underway between stations (on average, half the distance). This value is now used in the formulas above by simply adjusting the integration limits from "3" to "$3 + d_n$". This procedure is justified if the integrand does not vary strongly over the range of allowed values for the driving time margin. For the purpose of obtaining a rough estimate of the correction term, this is the case.

Delay may Persist After First Measurement
Our formalism measures the punctuality effects at the first measurement point (the next node). In fact, an effect may persist at later nodes, if the delay is not absorbed into the halting time at the node. This effect has been estimated at the second node by determining the average halting time margin hn from the timetables and adjusting the integration limits from "3" to "$3 + d_n^{(1)} + h_n + d_n^{(2)}$". Here $d_n^{(1)}$ is the average driving time margin from the sequencing point to the first node and $d_n^{(2)}$ the average driving time margin between nodes. They are added together with hn because at the second node, driving time margins of both the first and second stretch and one halting (at the first node) apply. Effects at later nodes can be computed in the same manner, but were found to be negligible.

Again, this procedure is justified if the integrand does not vary strongly over the range of allowed values for the halting time margin. Actually, this is not the case. Since, due to the exponential falloff of the delay distribution, trains with large halting time margins do not contribute strongly to the results, we have dealt with this problem by using an ad-hoc cut-off value on the halting time margins, considering only trains with small margins for our rough estimate of this correction term. Essentially, this means we consider trains passing through a station but not those at end nodes.

Delay may Cause Other Delays
We have not attempted to obtain a quantitative estimate for cascade effects. Operational data suggest strongly these effects are small.

3 Inputs to the Calculation

As mentioned earlier, contact with actual data is made through the if-then sheets. From these if-then sheets we obtain the number of potential resequencing decisions $A(d_c)$ on a day. Because of the corrections terms in Section 2.3, a distinction has to be made between resequencing decisions made between stations (open track) and at stations. Note: the majority of the resequencing decisions on stations concern departure situations.

Table 1. Number of potential resequencing decisions

Location \ d_c :	1	2	3	4	5	6	7	8	9	10	11	12	13
Open Track	87	336	497	594	752	454	539	695	461	306	260	181	160
Stations	250	714	1319	1070	1163	619	745	942	577	574	627	307	315

There are 5000 trains each day. The punctuality is measured at large stations (nodes), the average number of nodes for a train is 2.2, giving a total number of 11000 punctuality measurement points MS.

The value $d_{\text{av}} = 1.8$ minutes for the negative exponential delay distribution reproduces the 2002 punctuality of 82% for the Dutch railways. The conflict time or minimal headway (H) for insertion points is, according to headway calculations set to 90 seconds on the open track. For departure situations, values range from 97 to 133 seconds. We use 110 seconds.

The inaccuracy of the dispatching function in the sequence decision is set to $\sigma = 2$ minutes. This inaccuracy is composed of:

- rounding errors: in the decision process, dispatchers use three different delays, independently rounded to whole minutes;
- inaccuracy in the operating times themselves (30 - 60 sec);
- inaccuracy in the prediction of the delay at the insertion point;
- inaccuracy in the predicted travel time of the train after the insertion point;
- inaccuracy in the human dispatchers' decision making process.

Of these, the first two items are objectively known, for the others we have to rely on expert estimates. We are confident that the true standard deviation of the distribution will have a value between 1.5 and 3 minutes. More accurate determinations are feasible but have not been part of this investigation. We have used a shift (Formula 3) of 1 minute.

The parameters in the correction terms mentioned in Section 2.3 have been set as follows. For d_n, the driving time margin, i.e., the amount a train on average can reduce its delay between nodes, 1.7 minutes is used. The average halting time margin hn, after applying a cut-off value of 4 minutes to the data, was found to be 0.7 minutes. These values have been extracted from the timetable planning in the Netherlands.

In summary, unless stated otherwise we have used the following values for the parameters in our calculation:

Table 2. Default parameter values

Parameter	Symbol	Default value
Average delay	d_{ev}	1.8 minutes
Headway	H	90 seconds (open track)
		110 seconds (departures)
Dispatching error	Σ	2 minutes
Number of potential resequencing decisions	$A(d_c)$	Refer to Table 1
Average node-node driving time margin	d_n	1.7 minutes
Average halting time margin after cut-off	h_n	0.7 minutes
Cut–off value	h_{cut}	4 minutes
Shift	s	1 minute

4 Results

With the formulas from Section 2 and the default parameter values from Table 2 the effects of inaccuracy in the resequencing decisions can be computed. Table 3 gives the overall punctuality effect, the chance of an incorrect decision per train and the average loss per train, both with and without the correction terms mentioned in Section 2.3.

From the discussion so far, it should be obvious that our calculations depend upon a number of parameters of which the values are not accurately known. We have therefore experimented extensively with our formalism, in order to get a feeling for the range of possible results. As an illustration of these efforts, in Figure 3 we present the dispunctionality effect in cases where all parameters but one have been set to their default values (refer to Table 2) and one is varied. On

Table 3. Overall results

	Without correction	With correction
Punctuality effect	4.16%	3.62%
Chance of error	15%	
Average loss	0.5 minutes	

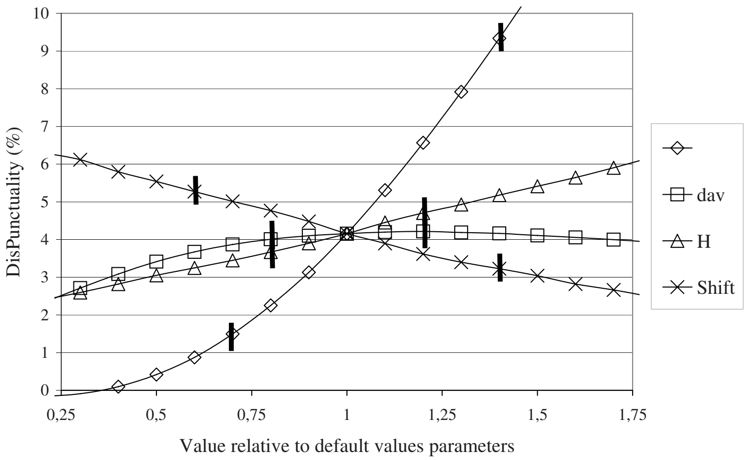

Fig. 3. Sensitivity analysis dispunctuality on input parameters

the horizontal axis the parameter value is given as a fraction of the default value. On the vertical axis, the resulting punctuality effect without correction terms is given. The vertical bars on the curves picture the range of plausible values for the parameter. Variations on the parameter values for d_n, h_n are similar to the variations on parameter s and not depicted in Figure 3. As to be expected, sensitivity is particularly high for the value of σ. Experimental determination of the control error distribution is highly desirable to narrow down our estimate. We will discuss the consequences of this (and similar) figure in the next section.

5 Discussion and Conclusions

The numbers given in Table 3 represent our best estimate of the effect of dispatching errors on punctuality. The numbers are unexpectedly large: around 4% of the observed dispunctuality stems from suboptimal resequencing. This is not in line with common opinion, which holds that suboptimal resequencing only contributes marginally to dispunctuality. Still, though many ingredients of our calculations are less than certain, our basic reasoning is so simple that it is hard to refute: the resequencing decision has an intrinsic accuracy of at least 1.5 minutes, there are many trains with delays between 1 and 3 minutes and many potential resequencing decisions (if-then rules) in this domain. Large effects are unavoidable. Note that the fact that the Dutch railway network is heavily used is crucial in this reasoning, since this ensures that many resequencing decisions have to be taken at delay values around 3 minutes, influencing punctuality. Our results are being corroborated by recent results of simulation study using the Combine Traffic Management System, which confront an “ideal” traffic management system (i.e. Combine) with data from practice at a specific location and find punctuality gains in the order of 5%.

As stated, our calculations were made within the context of an effort to construct a business case for investing in the dispatching control systems. For this purpose, we really need to know only that effects are large, say more than 0.5%. As our sensitivity analysis shows, varying the input parameters within the range of plausible values does not change this conclusion. The sensitivity analysis also makes us confident that our formalism, which relies on the use of average distributions, is reasonable.

Some other objections are possible.

- *Some inaccuracy is intrinsic to the process, therefore the "ideal decision making process" has no practical relevance.* Indeed, some errors are unavoidable. Especially the halting process at stations is unpredictable at the level of some tens of seconds. However, as Figure 3 shows, errors under 0.5 minutes do not contribute strongly to the overall result, while effects grow roughly linearly from $\sigma = 1.5$ to $\sigma = 3$ minutes. Therefore our conclusion holds.
- *Real sequencing is not done according to a simple summed delay criterion, other arguments are taken into account.* We have examined the if-then sheets, which have been prepared by hand by local experts for all single-delay situations, and found that our simple rule reproduces the if-then rule in very nearly all cases. Also the number of cases where a third train is relevant to the decision is not large. In an evaluation of a pilot of a conflict detection and resolution system based upon pairwise decisions, only very few exceptions were found.
- *The delay distribution (1) includes the effects, therefore cannot be input to the calculation.* This argument only holds merit if we find large effects, in which case our purpose is already met. Apart from that, we are very insensitive to the value of d_{av}. We have explicitly tested this point by varying dav to reproduce the actual punctuality including our effects. The predicted dispunctuality was almost constant.

The results are significantly larger than process experts expected. Why is that? We feel that for a long time the attention of human operators has been to maintain stable patterns (train sequences) close to the predefined timetable in order to maintain a workable mental image of the process. Optimizing at the level of plus or minus one/two minutes was neither possible, nor aimed for. As a result, the overall external goal of the railway sector, set at very definite punctuality levels in a very specific definition, a goal for which these relatively small errors are highly relevant, has only been aimed for in an indirect way and slipped out of focus.

In conclusion: our calculations were made within the context of an effort to construct a business case for investing in the dispatching control systems. It appears this investment would be an order of magnitude more effective than other punctuality oriented investments currently being suggested. In fact, even if our calculation overestimates the real effect by a factor of 2, which is within the realm of possibilities, the business case is still easily made.

Compared to the strategic (infrastructure planning) and tactical (timetable design) phase of railway operations, the operational phase has received relatively little attention from the mathematical community. Most work for the operational phase focuses on constructing real-time scheduling engines such as Marco and Combine. Experience suggests that for processes with are not under very precise control, as railway operations in the Netherlands neither are nor will be in the foreseeable future, an approach using very simple operational rules, which are well understood and easily amended by the operators, is to be preferred. Our calculations show that in the Netherlands, accurate implementation of one such rule may unlock significant gains in punctuality. We are aware that our formalism is quite basic. We challenge the mathematical community to improve upon our ideas, or indeed come up with an alternative, to get an approach to determine the effectiveness of specific simple dispatching rules.