Ex ante heuristic measures of schedule reliability

Malachy Carey*

Faculty of Business and Management, University of Ulster, Newtonabbey BT37 OQB, UK

Received 20 April 1998; received in revised form 9 November 1998; accepted 10 December 1998

Abstract

Measures of reliability and punctuality of scheduled public transport services are important in planning, management, operating and marketing of these services. Various methods can be used to measure reliability. Analytic methods are usually practical for only very simple structured systems. Simulation methods are very time consuming and require data which may not be available. As a result, the most widely used measures are ad hoc or heuristic. However, the assumptions and properties of these measures, and the relationships between them, are seldom discussed, hence we discuss them here. We consider existing measures, extensions of these, and new measures. For specificity, we use the example of train arrivals and departures at a train station: stations with several hundred trains per day and multiple platforms are common in many countries, for example throughout Europe. Some measures of reliability are based on the observed delays, hence can be used only after the event. However, we here focus on measures which can be used in advance, for example for estimating the reliability of proposed schedules or changes in schedules at the design stage. In this we distinguish between measures which require some information about probabilities of delay and those which do not. We also distinguish between exogenous delays, which are beyond the influence of the scheduler (delays due to problems in engineering or operations), and knock-on delays which are affected by schedule design: both types are of interest for schedule reliability, but the latter are of more interest for measuring schedule robustness. © 1999 Published by Elsevier Science Ltd. All rights reserved.

Keywords: Reliability measures; Transport planning; Trains

1. Introduction

Heuristic measures of reliability of scheduled public transport are widely used (for example, the percentages of services on-time, or more than 5, 10, etc., min late, or the mean lateness). To operators these measures are useful in planning, management, control, dispatching and marketing of these services. To users they are needed in making travel choices. To regulators they are needed to check if operators are delivering the promised or contracted quality of service. Operators (for
airlines, buses and trains) are now in many cases required by law or regulation to regularly publish heuristic measures of reliability or punctuality.

However, the assumptions and properties of these heuristic measures or indicators and the relationships between them are seldom discussed. Here we discuss these, considering existing measures that are used formally or informally in practice and some extensions of these and some new measures. When introducing each of these heuristics we give some intuitive motivation for it as a measure of reliability. Such motivation is important since without it these measures would not be used in practice. For specificity, we use the example of scheduled train arrivals and departures at a busy train station. Such stations with several hundred trains per day and multiple platforms are common in many countries, for example throughout Europe.

Transport schedules are usually developed and presented as fixed sets of times, for arrivals, departures, etc. But in practice schedules are operated under conditions involving on-the-day disturbances, delays, breakdowns, etc., which in turn cause further knock-on delays. Measuring the punctuality or reliability of such scheduled services after the event is relatively easy in principle. Most public transport operators maintain detailed records of punctuality and reliability, and are often required to do so by law or regulation. However, what is often of more interest to transport planners or managers is the ability to predict reliability in advance, especially for proposed new schedules or services or when considering service changes. The reliability of schedules may be tested in advance by conducting extensive simulations (for example, Carey and Carville, 1998; Hallowell and Harker, 1998). However, such simulations are very time consuming to design and run. Faster, even if less accurate, measures of reliability are often needed, and very crude measures are often used in practice.

For example, at the planning and design stage numerous variants of schedules are often considered and planners would like to have a quick way to compare the reliability of these. Also, throughout the year, operators are constantly revising and rescheduling in response to short run changes in demand, availability and operating conditions, and would like to have quick estimates of the effect on reliability of proposed schedule changes. The same is true for on-the-day rescheduling in response to unplanned delays. In each of these cases of short, medium and long term planning, it is usually not practicable to undertake a full scale simulation to evaluate every possible schedule change. Simple, quickly computed, heuristic measures or rules of thumb are used instead.

Heuristic measures of reliability are also useful for other management purposes. For example, if we measure the reliability of the schedules in the same way for different services or locations this can provide benchmarks for performance of operators and managers of these services or locations. Operators and managers already use ad hoc rules of thumb for reliability comparisons, for example the average amount of headway for certain services, or the number of services having less than a certain amount of headway.

In practice, ex ante heuristic measures or indicators of reliability may differ significantly from ex post measures. If the former systematically over or underestimates the latter, this may not be a serious problem since one can make a compensating correction. Alternatively, ex ante measures might underestimate the ex post measures for some locations or services and overestimate them for others. But even that may be useful if it draws attention to causes of differences in reliability. For example, differences may be due to local differences in management or equipment.

In considering reliability it is useful to consider the two types of delays, exogenous delays and knock-on delays. Exogenous delays are due to events such as breakdown or failure of equipment or infrastructure, delays in passenger boarding or alighting, lateness of operators or crews, etc.
Exogenous delays are generally not caused by the schedule. In contrast, knock-on delays are due to exogenous delays and the interdependence in the schedule. These exogenous delays frequently cause knock-on delays to other trains, which can in turn cause further knock-on delays. For example, if a train is late leaving a station platform this may delay the arrival of the next train scheduled to use the platform, which may in turn delay further trains. Or, if a train arrives late its scheduled platform may be already occupied, so that the train has to be sent to a different platform which may in turn delay trains scheduled for that platform.

Knock-on delays (but not exogenous delays) can be reduced by better scheduling, for example by giving more headway to trains which are prone to exogenous delays. Since scheduling can reduce knock-on delays but not exogenous delays, we can define a reliable schedule a’s one in which exogenous delays cause the least knock-on delays. We assume throughout that the schedule for which we wish to estimate reliability is a feasible schedule. A schedule is said to be infeasible if trains could not meet the schedule even in the absence of all unscheduled delays. For example, it is infeasible if two trains are scheduled to arrive at the same platform at the same time, or to pass each other on a single line, or if less than minimum required headways or dwell times are allowed.

There is surprisingly little published discussing ex ante measures of reliability, though there is a literature on other related aspects of transport reliability. For example, as already mentioned, Hallowell and Harker (1998) and Carey and Carville (1998) generate measures of reliability by simulating train running for given schedules, which is very time consuming. Chen and Harker (1990) estimate distributions of delays. Dejax and Bookbinder (1991) discuss ways in which a major transport operator (SNCF) can measure, monitor and market reliability. Allen et al. (1985) consider transit in an overall logistics context and include the mean and variance of trip times in the cost overall cost function. Carey (1994) examines the knock-on effects of delays in interconnected scheduled services. Carey (1998) considers inserting slack time in schedules to allow for variability and how this is affected by the behavioural response of transport personnel and users. There is a substantial literature on transport scheduling, dispatching and control and this is often implicitly or explicitly concerned with producing more reliable schedules, though the various measures of reliability are seldom discussed. There is some unpublished work on ex post measuring or assessing the reliability and performance of transport operations (e.g. Benwell and Black, 1984), or discussing the effects of (un)reliability, for example the effects on demand (Black and Trowiss, 1993). Also, of course, transport operators produce a large volume of data on actual punctuality and reliability, but this is all after the event, and is usually in internal unpublished reports.

In Section 2 we set out heuristic measures of reliability which assume we have information about the probability distributions of exogenous delays. In Section 3 we set out heuristic measures not using any probability information. Section 4 concludes, and in appendices we derive some properties of some of the main heuristic measures.

2. Heuristic measures of reliability, using probabilities

2.1. Notation

Before proceeding we introduce some notation. Consider trains arriving and departing from a single station and let:
\(t, t', t''\) denote three successive scheduled train movements (arrivals or departures) such that \(t\) is immediately before (earlier than) \(t'\), and \(t''\) is immediately after \(t'\).

\(a, d\) denote the actual arrival and departure times respectively of train \(t\).

\(m_t\) denote the time of the \(t\)th train movement, or movement time of the \(t\)th train.

Since much of the analysis and discussion is the same for train arrivals and departures, to avoid repetition we will often generically refer to train movements rather than arrivals or departures, hence \(m_t\) rather than \(a_t\) or \(d_t\). Also, though \(t\) refers to a train movement rather than a train, for short we often refer to \(t\) as a train. If a train arrives and departs in the schedule, both movements are listed separately so that the train is listed twice.

In Section 3 we divide the set of trains \(T\) into subsets \(\tilde{T}\) of different types and let \(S_{\tilde{T}}\) be the set of different types, so that \(t \in \tilde{T}\) and \(\tilde{T} \in S_{\tilde{T}}\).

2.1.1. Data

\(a^t, d^t, m^t\) are scheduled arrival, departure, or movement time respectively of train \(t\).

\(f_t(a^t), f_t(d^t), f_t(m^t)\) are the pdfs for the arrival, departure, or movement times of train \(t\). (We should put a subscript \(a, d, m\) on \(f_t\) but, as we do not use these together, omitting it causes no confusion.)

\(h^t_{st}\) is the scheduled headway between \(t\) and \(t'\).

\(h^t_{fr}\) is the minimum required headway.

\(h^t_{ef} = (h^t_{st} - h^t_{fr})\) is the free headway between trains \(t\) and \(t'\), that is, the maximum amount by which train \(t\) can be delayed without causing any knock-on delay to the next train \(t'\).

\(h = [h^t_{1,2}, h^t_{2,3}, \ldots, h^t_{T-1,T}]\).

2.1.2. Reliability measures computed from these data

\(r_{ht}(h^t_{ef})\) denotes a heuristic measure of reliability for train \(t\).

\(R_{h1}(h), R_{h2}(h), \ldots\) denote heuristic measures of reliability for the train schedule, based on using probability information. If these include a cost or importance weight for each train add a superscript \(w\), for example \(R^{w}_{h1}(h)\).

\(R_{ht}(h), R_{ht}(h), \ldots\) denote heuristic measures of reliability for the train schedule, not using probability information.

2.2. Using probabilities of delays

The percentage of services which are on-time, and the percentages more than 5, 10, etc., min late are often used as measure of reliability. These percentages are obtained from the observed frequency distribution of lateness. The percentages and frequency distributions can be taken as estimates of probabilities and probability distributions, but the latter terminology is not generally used by operators since the language of probability is not as well understood by the public. However, we will here refer to probabilities and probability distributions since this is simpler and more natural for discussion and analysis. As a measure of reliability we can use the probability \(P(.)\) that movement \(t'\) is on-time, that is, the probability that it does not move (arrive or depart) after its scheduled time \(m^t_{t'}\). This probability is given by \(\int_{m^t_{t'}}^{\infty}(pdf\ \ of\ \ movement\ \ time)\)dt. However, the “pdf of the movement time” \(m^t_{t'}\) is a combination of the movement time pdf in the absence of any knock-on delays [i.e. \(f_{t'}(.)\)], and the pdf of knock-on delay from the preceding train \(t\), which may in turn depend on knock-on from earlier trains, and so on. Defining or calculating the pdf
for this chain of knock-ons is extremely complex and in any case it is unlikely that all of the data are available.

In view of all this, we will set out some approximations to the “pdf of the movement time” of train \( t' \), first by ignoring knock-on delays, then by considering only knock-ons from the immediately preceding train, and finally by considering knock-ons from earlier trains.

2.2.1. Own delays

Ignoring knock-on delays, the pdf of the departure time of train \( t' \) is \( f_t(\tau) \).

\[
 r_{h_{t'}}(h_{t'}) = P(\text{train } t' \text{ does not suffer an own-delay})
 = P(m_t + h_{t'} \leq m_t') = \int_{-\infty}^{m_t'} f_t(\tau) d\tau.
\]  

(1)

2.2.2. One-stage knock-on delays

Even if there is no own-delay, train \( t' \) may suffer a knock-on delay from preceding trains. In practice most train conflicts involve only two consecutive neighbouring trains and indeed to try to ensure this, at least small amounts of ‘slack time’ or ‘recovery time’ are usually built into train schedules. In view of this, as an approximation we make the following simplifying assumptions, which we relax later below.

**Assumption 1.** Any knock-on delays to the movement (arrival or departure) of train \( t' \) are caused only by exogenous delays of the immediately preceding train movement \( t \), defined by the pdf \( f_t(\tau) \)—knock-on delays to \( t' \) from trains before train \( t \) are negligible.

**Assumption 2.** The sequence order of the trains is always maintained, no matter how long trains \( t \) or \( t' \) are delayed.

Then a heuristic measure of reliability for train \( t' \) is the probability that it cannot be knocked-on by the preceding train \( t \), which is the probability that train \( t \) keeps within its headway \( h_{t'} \), thus,

\[
 r_{h_{t'}}(h_{t'}) = P(m_t + h_{t'} \leq m_t') = \int_{-\infty}^{m_t' + h_{t'}} f_t(\tau) d\tau
\]  

(2)

This definition of reliability implies that the larger the headway \( h_{t'} \) the greater the reliability of train \( t' \) (the less likely it is to suffer knock-on delays). Also, \( \partial r_{h_{t'}}(h_{t'}) / \partial h_{t'} = f_t(m_t' + h_{t'}) \), hence reliability always increases with headway \( h_{t'} \), since \( f_t(\tau) \) and \( h_{t'} \) are always positive. Also, in practice the probability density \( f_t(\tau) \) eventually tails off (decreases) hence reliability \( r_{h_{t'}}(h_{t'}) \) eventually increases at a decreasing rate, which is intuitively reasonable.

2.2.3. Own delays plus one-stage knock-on delays

Now consider both the above types of delays: knock-on delay plus own delay. Train \( t' \) can move at any time \( m_{t'} \), only if the preceding train \( t \) has already moved at least \( h_{t'} \) min earlier,
where $h_{tr}'$ is the minimum required headway. The probability of this happening is $\int_{-\infty}^{m_r-h_{tr}'} f_i(\tau)d\tau$ hence the probability density for train $t'$ moving at any time $m_r$ is $f_{tr}\int_{-\infty}^{m_r-h_{tr}'} f_i(\tau)d\tau$. Hence the knocked-on pdf for the movement time of train $t'$ is,

$$
f^k_{tr}(m_r) = f_{tr}(m_r) \int_{-\infty}^{m_r-h_{tr}'} f_i(\tau)d\tau. \quad (3)
$$

Using this, the reliability (the probability of no delay to train $t'$) is formally the same as (1) but with $f_{tr}(.)$ replaced by $f^k_{tr}(.)$,

$$
r_{h_{tr}}(h_{tr}) = P(m_r \leq m_r') = \int_{-\infty}^{m_r'} f^k_{tr}(\tau)d\tau. \quad (4)
$$

### 2.2.4. Multi-stage knock-ons: relaxing Assumption 1

Now drop the assumption of a single-stage knock-on delay (Assumption 1) and consider multistage knock-on delays. Consider trains listed in chronological order of their scheduled movements (arrivals and departures), and suppose (Assumption 2) the sequence order of the train movements has to be maintained (i.e. train 1 before 2, 2 before 3, and so on). The pdfs for the movements of trains $1, \ldots, t, t'$, in the absence of any knock-on effects are $f_{i}(.), \ldots, f_{i}(.)$. Then by analogy with (3), the knocked-on pdf of train movement $t'$ is,

$$
f^k_{tr}(m_r) = f_{tr}(m_r) \int_{-\infty}^{m_r-h_{tr}'} f^k_{i}(\tau)d\tau. \quad (5)
$$

where $f^k_{i}(m_t)$ is the knocked-on pdf for the preceding train movement $t$, that is,

$$
f^k_{i}(m_t) = f_{i}(m_t) \int_{-\infty}^{m_t-h_{i-1}'} f^k_{i-1}(\tau)d\tau. \quad (6)
$$

and so on recursively backward to train 1, the first train in the sequence. Fortunately, to compute $f^k_{i}(.)$ we normally need consider at most the last few trains in the sequence since these are likely to have the largest effect on $f^k_{i}(.)$. If the free headway between a neighbouring pair of trains in the sequence is large, so that probability of knock-on delays between them is small, then any trains before this point can be ignored when computing $f^k_{i}(.)$.

More specifically, if the pdfs for neighbouring trains in the sequence (say trains $\tau$ and $\tau - 1$) have no overlap then the integral term in (5) will be equal to 1, in which case (5) reduces to $f^k_{i}(m_t) = f_{i}(m_t')$ so that there is no knock-on effect from train $\tau - 1$ to $\tau$. Similarly, if the pdfs for trains $\tau - 1$ and $\tau$ have little overlap then there is little knock-on effect from train $\tau - 1$ to $\tau$. Trains tend to have less and less knock-on effect on later and later trains in the sequence.

Introducing these multi-stage knock-on delays (5) into the reliability measure (4), gives,

$$
r_{h_{tr}}(h_{tr}) = \int_{-\infty}^{m_r'} f^k_{tr}(\tau)d\tau, \text{ with } f^k_{tr}(.) \text{ from (5) rather than from (3).} \quad (7)
$$
Actually, the situation is more complex than above. When computing knock-on effects from preceding train movements, some of the pdfs should be truncated and/or shifted. For example, trains require time at stops for boarding, alighting, checking, failures, etc., hence the time at which they are “ready-to-depart” is a random variable with pdf say \( f_r(.) \). However, even if a train is ready to depart, it is usually not permitted to depart before its scheduled departure time. Let \( f(.) \) be the pdf of its permitted departure time. To obtain \( f_t(.) \), take the left hand tail of \( f_r(.) \), prior to the scheduled departure time, and treat it as a point mass at the scheduled departure time in \( f_t(.) \). The right hand tails of both pdfs (beyond the scheduled departure time) are the same. This is discussed in more detail in Carey (1994) and Carey and Kwieciński (1995). Note that \( f_t(.) \) is the pdf of permitted departure time in the absence of knock-ons from other trains. This is then used in (3) or (5) above to obtain the pdf \( f_t(.)^k \) including knock-ons.

2.2.5. Probability (or percentage) of trains more than 5, 10, etc., min late

Above we considered the probability of a service being on-time. On-time is equivalent to being less than or equal to \( x = 0 \) min late. We can easily extend the discussion and results to the probability of being less than or equal to \( x = 5 \) min late, \( x = 10 \) min late, and so on. To do that we simply add \( x \) to the upper limit of integration in Eqs. (1)–(7). For example, in Eq. (1) \( m_r + x \), in Eq. (2) \( m_t + h_t \) becomes \( m_t + h_t + x \).

2.2.6. Aggregate reliability (schedule reliability)

As a heuristic measure of aggregate reliability for the set of trains \( T \), we take the mean of the individual train reliabilities \( r_{h1,.}(h_{tr}) \), thus,

\[
R_{h1}(h) = \left( \sum_{r=1}^{T} r_{h1,.}(h_{tr}) \right) / T = \text{mean of the probabilities that trains} \quad t = 1, \ldots, T, \text{do not suffer knock-on delays,}
\]

where \( r_{h1,.}(h_{tr}) \) is given by (1), (2), (4) or (7).

Since \( R_{h1}(h) \) is a mean of probabilities it lies between 0 and 1. It will be 1 only if there is no possibility any of delays, of the types considered in (1)–(7). It will be 0 only if it is certain that every movement will suffer some delay, of the types considered in (1)–(7).

2.2.7. Arrivals, departures, dwell times, etc.

For convenience in the discussion above we referred to train movements, which can be either arrivals or departures. The expressions for \( r_{h1,.}(h_{tr}) \) and \( R_{h1}(h) \) can be rewritten explicitly listing all types of train conflict: train \( t \) arriving and \( t' \) departing, \( t \) departing and \( t' \) departing, etc. Also, if we wish to consider only some types of conflict then we need only change the above expressions accordingly. For example, if we wish to consider only delays to arrivals caused by movements of preceding trains then in (8) change \( T \) to \( T^v \) where \( T^v \) is the set of arriving trains. Conversely, if we wish to consider only delays to train movements caused by arrivals of preceding trains \( t \) then in (1), (2), (4) and (7) change \( m_t \)'s in the integral limit to \( a_t^r \).

2.2.8. Considering only subsets of services.

We may be interested only in conflicts caused by or to particular subsets of transport services, for example trains using the same platforms. Let \( T_p \) be...
the subset of the trains $T$, in chronological order, that belong to the same service $p \in P$, and let $|T_p|$ be the number of trains in this set. Then a measure of reliability including only delays to these trains $t' \in T_p$ is obtained by replacing $\sum_{t'=1}^T$ in (8) with $\sum_{p \in P \cap t' \in T_p}$.

2.3. Using expected delay instead of probability of delay

The above measures of reliability are all based on the probability of knock-on delays. We can easily adapt this to instead use the expected size of the knock-on delays. As above, we consider own-delays, one-stage knock-on delays and multi-stage knock-on delays separately.

2.3.1. One-stage knock-on delays

Knock-on delay to train $t'$ occurs if the preceding train $t$ is delayed by more than $h_{tt'}$, hence, expected knock-on delay to train

$$t' = E(m_t - (m_t^i + h_{tt'})) = \int_{m_t^i + h_{tt'}}^{\infty} \tau f_i(\tau) d\tau. \quad (9)$$

To put this on a scale of 0 to 1, which is convenient for reliability measures, we can express it as a fraction of the expected delay which would obtain if the headway was zero, that is,

$$E(m_t - m_t^i) = \int_{m_t^i}^{\infty} \tau f_i(\tau) d\tau.$$ Then a measure of unreliability on a scale 0 to 1 is,

$$ur_{h, t}(h_{tt'}) = \left[ \int_{m_t^i + h_{tt'}}^{\infty} \tau f_i(\tau) d\tau \right] / \left[ \int_{m_t^i}^{\infty} \tau f_i(\tau) d\tau \right].$$

This measure decreases the more reliable the schedule becomes. To obtain a measure which increases with reliability we subtract from 1, thus,

$$r_{h, t}(h_{tt'}) = 1 - \left[ \int_{m_t^i + h_{tt'}}^{\infty} \tau f_i(\tau) d\tau \right] / \left[ \int_{m_t^i}^{\infty} \tau f_i(\tau) d\tau \right]. \quad (10)$$

Alternatively, we can construct a measure of reliability from (9) by taking its inverse thus,

$$r_{h, t}(h_{tt'}) = 1 / \left[ \int_{m_t^i + h_{tt'}}^{\infty} \tau f_i(\tau) d\tau \right]. \quad (11)$$

This increases as expected delay decreases but is on a 0 to $+\infty$ scale.

To obtain measures of reliability for the whole schedule, take the mean of any of the above measures over all trains thus,

$$R_{h}(h) = \left( \sum_{t=1}^T r_{h, t}(h_{tt'}) \right) / T.$$ 

(12)
or alternatively,

\[ R^*_h(h) = \left( \sum_{r=1}^{T} r_{h,\tau}(h_\tau) \right) / T \]  

both of which lie between 0 and 1. It is of interest to compare the reliability measure \( R_{h1}(h) \) which is based on the probabilities of delays and the measures \( R_{h2}(h) \) and \( R^*_h(h) \) which are based on expected delays.

**Proposition 1.** Let \( f_i(m_t^s + \tau r) = f(m_t^s + \tau) \) for all trains \( t \), i.e., the train delay pdfs are the same for all train movements. Also, let the reliability measures \( R_{h1}(h) \), \( R_{h2}(h) \) and \( R^*_h(h) \) be used to rank any set of schedules in increasing (or decreasing) order of reliability.

Then all three measures \( (R_{h1}(h), R_{h2}(h) \) and \( R^*_h(h) \)) yield the same rank order for the set of schedules.

**Proof.**

1. If \( \int_{-\infty}^{m_t^s + \tau r} f_i(\tau) d\tau \) (from \( R_{h1}(h) \)) ranks higher for train \( t_1 \) than for train \( t_2 \), then \( \int_{m_t^s + \tau r}^{\infty} f_i(\tau) d\tau \) will rank lower for train \( t_1 \) than for train \( t_2 \), hence \( \int_{m_t^s + \tau r}^{\infty} f_i(\tau) d\tau \) (from \( R_{h2}(h) \)) will rank lower for train \( t_1 \) than for train \( t_2 \). Dividing the latter by \( \int_{m_t^s}^{\infty} f_i(\tau) d\tau \) does not change the ranking, since \( \int_{m_t^s}^{\infty} f_i(\tau) d\tau \) is the same for all \( t \). Summing the terms in \( R_{h2}(h) \) does not change the ranking, and subtracting these fractions from 1 reverses the ranking. Hence the final ranking from \( R_{h2}(h) \) is the same as from \( R_{h1}(h) \).

2. Now consider \( R_{h2}(h) \). From (i), if \( \int_{-\infty}^{m_t^s + \tau r} f_i(\tau) d\tau \) ranks higher for train \( t_1 \) than for train \( t_2 \), then \( \int_{m_t^s + \tau r}^{\infty} f_i(\tau) d\tau \) will rank lower for train \( t_1 \) than for train \( t_2 \). Summing the terms in \( R_{h2}(h) \) does not change the ranking, and taking the inverse reverses the ranking. Hence the final ranking from \( R_{h2}(h) \) is the same as from \( R_{h1}(h) \). \( \square \)

**2.3.2. Own delays and multi-stage knock-on delays**

The above reliability measures based on one-stage knock-on delays can be extended to include own-delays and multi-stage knock-on delays, in much the same way as in Section 2.1 above. Own-delay to train \( t' \) occurs if its movement time \( m_t \) is later than its scheduled movement time \( m'_t \), hence,

\[ \text{expected own-delay of train } t' = E(m_t - m'_t) = \int_{m'_t}^{\infty} f_{t'}(\tau) d\tau. \]  

(14)

The pdf of own-delay plus one-stage knock-on delay for train \( t' \) is \( f^{k}_{t'}(\cdot) \), from (3), and using this,

\[ \text{expected(own + knock-on)delay to train } t' = E(m_t - m'_t) = \int_{m'_t}^{\infty} f^{k}_{t'}(\tau) d\tau. \]  

(15)
Expected own-delay plus multi-stage delay is,

\[(14) \text{ with } f_t(.) \text{ replaced by } f_t^k(.) \text{ from } (5).\] (16)

The expected delay measures (14), (15) and (16) can all be used to construct corresponding measures of reliability as in (10) and (11) above.

2.4. Different weights, importance or cost for different services, locations, times of day, etc

Train operators usually consider punctuality or delay to have a different weight, importance or cost for different trains. The weights are usually based on train speeds, stopping patterns, expected number of passengers, fare levels, expected revenue, etc. The weight or importance attached to train punctuality or delay may also differ by time of day or by the platform the train is sent to. However, we can assume that the time of day and the platform for each train is already decided and included in the train weight, hence we here need only refer to the weight for each train.

We can include the weight for each train in the above heuristic measures of reliability. Instead of using \( r_{h,i}(h_{t-1,i}) \) as the measure of reliability of train \( t \), apply a weight \( w_t \) for each train, so that the heuristic measure of reliability \( R_{h}^w(h) \) becomes,

\[ R_{h}^w(h) = \frac{1}{T} \sum_{t=1}^{T} w_t r_{h,i}(h_{t-1,i}). \] (17)

To ensure that this still lies between 0 and 1 we should scale the weights \( w_t \) so that they sum to 1, i.e. divide the given weights \( w_t \) by \( W = \Sigma w_t \).

More generally, if more data are available we can let the train weighting function be \( w_t(h_{t-1,i})r_{h,i}(h_{t-1,i}) \) or \( w_t(h_{t-1,i}) \) so that the heuristic measure of reliability becomes,

\[ R_{h}^w(h) = \frac{1}{T} \sum_{t=1}^{T} w_t r_{h,i}(h_{t-1,i}). \] (18)

However, the simplest weights \( (w_t) \) are usually sufficient, given the data likely to be available. The train weights used by train operators are usually based on train type, say 10 for intercity express 8 for ordinary intercity, 7 for local fast trains, etc. We can use these as the weights \( w_t \) unless more suitable data are available. Introducing these train weights in the measures of reliability \( R_{h1}(h), R_{h2}(h), \) etc., we denote the new measures by \( R_{h1}^w(h), R_{h2}^w(h), \) etc.

2.5. Train rescheduling or train order swapping: relaxing Assumption 2

We now relax Assumption 2 from Section 2.2.2 above. In defining the above expected delay we assumed (Assumption 2) that if the movement of train \( t' \) is delayed by the movement (arrival or departure) of train \( t \) ahead of it, then train \( t' \) has no option but to wait indefinitely on train \( t \). However, this overestimates the delay to train \( t' \). In practice part of the job of train controllers
and dispatchers is to intervene when delays occur and to change the schedule so as to reduce knock-on delays. If the knock-on delay to train $t'$ is more than say $x_{tt'}$ min, dispatchers may swap the order of the trains (let $t'$ go before $t$), or send train $t'$ to a different platform, or take some other action to avoid further knock-on delay to train $t'$.

Introducing this decision rule has no effect on the measures of reliability $R_{h1(h)}$, etc., which are based on the probability of not incurring knock-on delay, since in that case there is no need to swap the train order. However, introducing this decision rule does affect the measures of reliability $R_{h2(h)}$ based on the expected size of knock-on delays. To compute the expected delay to train $t'$ we now consider only delays up to $x_{tt'}$, so that the expected knock-on delay to train $t'$ is now,

$$E[(m_t^i + h_{tt'} + x_{tt'}) - (m_t^i + h_{tt'})] = \int_{m_t^i + h_{tt'}}^{m_t^i + h_{tt'} + x_{tt'}} f_i(\tau)d\tau.$$  

Then the reliability measure $R_{h1(h)}$ becomes,

$$R_{h1(h)} = \left( \sum_{r=1}^{T} r_{h1,r}(h_{tt'}) \right) / T = (1/T) \int_{m_t^i + h_{tt'}}^{m_t^i + h_{tt'} + x_{tt'}} f_i(\tau)d\tau.  \quad (19)$$

In the earlier measures of reliability, knock-on delays to train $t'$ are caused only by trains preceding $t$. But now if train $t'$ is delayed beyond a certain point $(m_t^i + h_{tt'} + x_{tt'})$ the next scheduled train $t''$ will be allowed to go before it. This may cause a further headway delay $h_{tt''}$, to train $t'$, but on the other hand, if $t'$ delays $t''$ then one or both may be switched to different platforms where they no longer conflict. We do not pursue this here.

The cutoff time $x_{tt'}$ used by controllers for swapping train order may be anywhere between 0 and a large number, depending on the relative importance of the trains, the tightness of the schedule, the availability of other platforms, etc. If values of $x_{tt'}$ used by controllers are not known, we suggest for example the following heuristic. Let $w_t$ and $w_{t'}$ denote the importance weights of trains $t$ and $t'$, respectively. The headway required between $t$ and $t'$ if we swap their order is $h_{tt'}$. Then choose the swap over point $x_{tt'}$, so as to equate the weighted delays to the two trains at the swap over point. That is, set $x_{tt'}w_{t'} = h_{tt'}w_t$, hence set,

$$x_{tt'} = h_{tt'}w_t/w_{t'}, \quad (20)$$

so that if the trains are of equal importance $x_{tt'} = h_{tt'}$.

To see how the heuristic rule $x_{tt'} = h_{tt'}$ works, consider two problem cases. The first is when swapping the order of trains $t$ and $t'$ would put a slow train $t'$ ahead of a fast train $t$ on the same out line from a station. If it is a long way to the next passing point on the line the fast train could be stuck behind the slow train for a long time. However, this is automatically prevented by the above heuristic rule, $x_{tt'} = h_{tt'}w_t/w_{t'}$. If train $t$ is faster than $t'$ then the minimum headway $h_{tt'}$ required between them can be set just large enough to ensure that the fast train $t$ does not catch up with the slow train $t'$ until their next potential passing point. If the trains are of equal importance ($w_t = w_{t'}$) the rule reduces to $x_{tt'} = h_{tt'}$, that is, let the slow train $t'$ go first only if it can get to the next passing point just as the fast train catches up with it. If the fast train $t$ is less important
(i.e. \( w_t < w_{t'} \)), the rule implies swap sooner, so that the slow train can depart sooner, and if the fast train \( t \) is more important, the rule implies delay swapping it.

The second problem case arises if the trains in question \((t \text{ and } t')\) are going to the same platform. This turns out to be formally very similar to the above case. If the trains are swapped, so that \( t' \) goes to the platform first, then it might seem that train \( t \) could wait a long time while train \( t' \) goes to the platform, completes its dwell time and departs. This could be a long wait: for example, in Britain some trains have a scheduled dwell times of 20, 30 or more min between long intercity trips. However, again this is automatically taken care of by the above heuristic rule, \( x_{tt'} = h_{tt'} w_t / w_{t'} \).

If the dwell time of train \( t' \) is large this is included in the required minimum headway \( h_{tt'} \), and if \( h_{tt'} \) is large \( x_{tt'} \) will be similarly large. If the trains are of equal importance \((w_t = w_{t'})\) the rule reduces to \( x_{tt'} = h_{tt'} \), that is, let train \( t' \) go to the platform first, only if it can complete its dwell time and depart just before the other train arrives. If the train \( t \) scheduled to go to the platform first is less important, then the rule implies swap sooner, so as to allow train \( t' \) to get to the platform sooner. And if train \( t \) is more important the rule implies delay swapping it with train \( t' \).

We suspect that both types of heuristic measures of reliability (based on probability of delay, and based on expected values of delay) are likely to be reasonably good when the train station is not very congested. However, if it is heavily congested we suspect that the latter type of measure will be better, since it takes account of what happens when trains are delayed beyond their scheduled headways.

2.5.1. Some extensions and limitations

In the above heuristic measure of reliability \( R_{h3}(h) \) we introduced train rescheduling or train order swapping. However, we do not explicitly consider whether such trains are swapped to different platforms or continue to use their scheduled platforms. To consider this in detail would take us into detailed rescheduling, for which we would need to know which platforms are currently free, how long they will be free, etc. We avoid this here. By not modelling the details of swapping trains to other platforms we may obtain an underestimate or overestimate of the overall reliability of the schedule. However, this may not be important for our purposes. We wish to use these measures of reliability to compare schedules and for such comparisons it is important only to get the relative magnitudes correct—if they all underestimate or overestimate this need not affect their rank order.

3. Heuristic measures of reliability, not using probabilities

Most transport systems have at least some information available from which at least rough probability distributions for arrival or departure delays could be constructed. For example, there is usually some information as to the maximum delays, average delays, or approximately what fraction of delays are from 0 to 5 min, or 5 to 10, 10 to 20, etc., min.

However, train planners may prefer not to use this past information for several reasons.

1. In measuring punctuality we include both exogenous delays and knock-on delays. However, in measuring reliability we wish to take exogenous delays as given and consider the knock-on delays which these cause. For this we would ideally need separate data on exogenous
delays and knock-on delays, but in the available data these may not be separable. Some operators (for example the British train operating companies, and formerly British Rail) try to keep track of these two types of delays separately, but are not completely successful.

2. What matters to decision makers are future rather than past exogenous delays. These may differ from past exogenous delays in ways that are difficult to predict, due to changes in the causes of the delays (reliability of rolling stock, operating practice, etc.)

3. The exogenous delays at a station include knock-on delays from earlier stations and these depend on the schedule, and thus may be difficult to predict when the schedule changes.

4. Heuristic “rules of thumb” or targets not involving probabilities are more familiar to most transport planners, and indeed to most people. For example, to reduce train conflicts and knock-on delays, train planners often specify desirable headways for each train type (in addition to their required minimum headways), and treat these as a measure of reliability. In view of this, we construct some measures of reliability not involving probabilities. We define most of these so that they lie between 0 and 1, with larger values indicating higher reliability.

Despite the above, and despite not using probability explicitly in the measures of reliability, we will introduce and motivate each of the reliability measures below by considering whether an increase in the measure is likely to ensure that the probability of knock-on delays will decrease. Some of the measures of reliability set out below are based on the assumption that reliability is higher if (weighted) headways, especially smaller headways, are made less unequal for all trains, or all trains of the same type. Is this justified? That is, if headways are made nearer to equal does that ensure that the probability of knock-on delays will decrease? The answer is yes if the pdfs of arrival or departure times are downward sloping for delayed trains, which is likely in practice. Proposition A2(i) states that for any pair of neighbouring trains 1 and 2 the probability of knock-on delay is at a minimum when

\[ w_1 f_1(h_1) = w_2 f_2(h_2), \]

where \( w_t \) is the importance weight for train \( t \) and \( f_t(h_t) \) is the pdf for the arrival or departure of train \( t \). We can assume that for trains of the same type \( w_1 = w_2 \) and \( f_1(h_1) = f_2(h_2) \) only when \( h_1 = h_2 \). Hence for trains of the same type, the probability of knock-on delay is at a minimum when the headways are equal.

This argument that equality of headways minimizes knock-on delays can be stated intuitively as follows (see Propositions A1 and A2). If the total amount of headway available is fixed, and we give 1 min more headway to one train we must take 1 min from another. If the probability that the first train arrives in this minute \([f_1(h_1)]\) is greater than the probability that the second train arrives in this minute \([f_2(h_2)]\), then the swap is worthwhile, otherwise it is not. That is, if \( f_1(h_1) > f_2(h_2) \) then swap a minute of headway, and keep on swapping headway until \( f_1(h_1) = f_2(h_2) \). More generally, if trains 1 and 2 have importance weights \( w_1 \) and \( w_2 \) then, to minimize the probability of knock-on delays, headways should be swapped between trains until \( w_1 f_1(h_1) = w_2 f_2(h_2) \).

In some subsections below we divide the \( t = 1, \ldots, T \) trains into subsets \( \tilde{T} \) of different types and let \( S_{\tilde{T}} \) be the set of different types, so that \( t \in \tilde{T} \) and \( \tilde{T} \in S_{\tilde{T}} \).

3.1. Minimum headway targets or percentiles

As noted above, making headways more equal for all trains of the same type reduces the probability of knock-on delays. However, it is often not possible to make headways equal within
a train type. For example, due to the pattern of travel demand, headways may be short in peak periods and long at other times of day. If we cannot make headways equal then we note that the largest improvements in reliability (reduction in knock-on delays) usually come from increasing the shortest headways. (If the delay time pdfs slope downward then the benefit from each increase in headway diminishes as the headway increases.) In that case, to improve reliability we should make the smallest headways as large as possible. Hence a natural measure of reliability is, for each train type \(^{\sim}T\),

\[ R_{H1, \sim T}(h) = \text{the 5th, 10th, etc., percentiles of distribution of headways for trains type } \sim T, \]

or, \[ R_{H2, \sim T}(h) = \text{fraction or percentage of trains } t \in \sim T \text{ which have a headway larger than } \sim h_T, \]

where \(\sim h_T\) is some target headway or desired headway. For example, planners may desire say an extra 10 min of headway for intercity trains, but are happy with only 2 min extra between trains until headway for local trains. To focus on short headways, \(\sim h_T\) should be set so to be achievable for most trains of type \(\sim T\).

The second measure has the advantage that it relates headways to a suitable target \(\sim h_T\) for each train type, hence can yield reliability numbers \(R_{H1, \sim T}\) which are comparable or similar for very different types of trains. Because of that it is more useful when comparing across train types, hence more useful for an overall measure of reliability for the whole schedule, that is, \(R_{H2}(h) = \text{fraction or percentage of all trains which have a headway larger than their corresponding } \sim h_T.\)

3.2. (Weighted) headway spread

The above measure uses only the lower tail of the distribution of headways. Here we consider measures which take account of all headways. We argued above that the probability of knock-on delay decreases as we make the headways “more equal” for trains of the same type. For a given sum of the headways, i.e. \(\sum_{i \in \sim T} h_i = \sim H_T\), measures of headway dispersion or spread can be used as measures of headway inequality. For example, the range, variance (var), standard deviation (s.d.) or mean absolute deviation (m.a.d.) of the headways can all be used as measures of inequality of the headways, hence can be used here as measures of unreliability.

To compare reliability of different schedules, it is useful or essential to put these reliability measures on a common scale independent of measurement units. To put them on a 0–1 scale we divide by the (maximum–minimum) possible value of the measure of spread being used. All of the above measures (range, s.d., m.a.d., var) are zero when the headways are all equal (no spread) and all are at their maximum when all the available headway \(\sum_{i \in \sim T} h_i = \sim H_T\) is given to one train and no headway to the rest. This is also true for all even numbered higher moments \(\sqrt[4]{\sum_{i \in \sim T}(h_i)^4/\sim T}, \sqrt[6]{\sum_{i \in \sim T}(h_i)^6/\sim T}, \) etc.

Let \(s\) denote a measure of headway spread or inequality and hence unreliability. Converting to a 0–1 scale, unreliability \(= (s - s(\text{min}))/s(\text{max}) - s(\text{min})) = s/s(\text{max})\) since \(s(\text{min}) = 0.\) Hence reliability on a 0–1 scale is,

\[ R_{H2, \sim T} = 1 - s/s(\text{max}). \]  

If different train types \(\sim T\) have different weights or importance \(w_{\sim T}\) we can take the weighted average of their reliabilities as a measure of overall schedule reliability, thus,
\[ R_{H^2}(h_T) = \sum_{\hat{\tau} \in S_{\hat{\tau}}} w_{\hat{\tau}} R_{H^2, \hat{\tau}} \]

where the weights \( w_{\hat{\tau}} \) are scaled to sum to 1.

It can be shown that \( s(\text{max}) \) for various measures of headway spread \( s \) are as follows.

<table>
<thead>
<tr>
<th>Measure</th>
<th>( s(\text{max}) ), maximum possible value of ( s )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>( H_{\hat{T}} )</td>
</tr>
<tr>
<td>Variance (var)</td>
<td>( (H_{\hat{T}}^2/n^2)(n-1) )</td>
</tr>
<tr>
<td>Standard deviation (s.d.)</td>
<td>( (H_{\hat{T}}/n)\sqrt{n-1} )</td>
</tr>
<tr>
<td>Mean absolute deviation (m.a.d.)</td>
<td>( 2(H_{\hat{T}}/n^2)(n-1) )</td>
</tr>
</tbody>
</table>

The range is too easily affected by an extreme value and the variance is in squared units hence we suggest using the s.d. or m.a.d. in a normalised measure of reliability. Using these we obtain, on a 0–1 scale,

\[ R_{H^2, \hat{\tau}} = 1 - \frac{(\text{s.d. of headways})}{(H_{\hat{T}}/n)\sqrt{n-1}} \]  \hspace{1cm} (22)

or

\[ R_{H^2, \hat{\tau}} = 1 - \frac{\text{(m.a.d. of headways)}}{2(H_{\hat{T}}/n^2)(n-1)}. \] \hspace{1cm} (23)

If we use these measures to rank the reliability of train schedules which have the same number of trains and the same total headways \( H_{\hat{T}} \) available, then the denominator in \( R_{H^2, \hat{\tau}} \) is the same for each schedule hence does not affect the ranking. However, in practice it is usually the sum of the headways over all train types which is fixed, or approximately fixed. Within that fixed total headway, operators may compare alternative schedules in which some train types have more total headway \( H_{\hat{T}} \) and other train types have less. In that case, when comparing schedules it is important to know whether the above measures of reliability are still intuitively reasonable. This can be answered as follows.

**Proposition 2.** Suppose we keep the number of trains fixed but increase or decrease the total headway, allocating the change in headway among the trains in proportion to their existing headway. Then the numerical values of above measures of reliability \( R_{H^2, \hat{\tau}} \) do not change.

**Proof.** To implement the change in headway multiply the headway for each train by a constant, say \( k \), so that \( h_{\tau \hat{\tau}} \) and \( H_{\hat{T}} \) become \( kh_{\tau \hat{\tau}} \) and \( kH_{\hat{T}} \), respectively. In \( R_{H^2, \hat{\tau}} \) the \( k \) appears in the numerator and denominator and hence cancels out. \( \square \)

It follows that the above normalisation of s.d. and m.a.d. (scaling between their largest and smallest possible values) is crucial if we wish to use these measures to compare schedules which
have different total headways $H_{\tilde{T}}$. If we did not scale in this way then an increase in $H_{\tilde{T}}$ would increase s.d. and m.a.d., hence increase the measure of unreliability $R_{H2,\tilde{T}}$, which is undesirable since increasing $H_{\tilde{T}}$ allows more headway between all trains, which intuitively should not increase unreliability.

In Appendix 2 (Propositions A3 and A4) we consider changing the number $n$ of trains of type $\tilde{T}$ and show that this changes the above reliability measures in the right direction, though the outcome is not as simple as in Proposition 2 above. A desirable property for a heuristic measure of reliability is that even if we scale up the number of trains and the headways in the same proportion (and keep spread measures m.a.d. or s.d. unchanged), then the reliability value $R_{H2,\tilde{T}}$ would not change. Propositions A3 and A4 show that the reliability measures (22) and (23) do not have this property. In contrast, we can easily show (Proposition A5) that the following measures do have this desirable property.

\[
R_{H2,\tilde{T}} = 1 - \frac{\text{(s.d. of headways)}}{(H_{\tilde{T}}/n)}
\]

or

\[
R_{H2,\tilde{T}} = 1 - \frac{\text{(m.a.d. of headways)}}{2(H_{\tilde{T}}/n)}.
\]

Because of this (24) and (25) appear preferable as measures of reliability, especially when comparing schedules which have different numbers of trains.

### 3.3. Weighted average headway

Increasing the headways reduces knock-on delays, but if the total amount of headway available $\sum_{t=1}^{T}h_{t-1,t} = H$ is given, this sum is no use as a measure of reliability. However, as an ad hoc measure of reliability planners sometimes place different weights or importance on headways for different train types: let $w_t$ denote the importance of reliability for train $t$. For example, each extra minute of headway may be considered twice as important for one train type as for another. In that case the objective is implicitly to maximise the weighted sum $\sum_{t=1}^{T}w_{t}h_{t-1,t}$, hence this sum can be used as a measure of reliability,

\[
R_{H3}(h) = \sum_{t=1}^{T}w_{t}h_{t-1,t}.
\]

If the weights are scaled to sum up to 1, this is the weighted mean headway. If the weights $w_t$ are all equal then $R_{H3}(h)$ reduces to $H$. If $w_t$ is the same for all trains $t \in \tilde{T}$, then $R_{H3}(h)$ can be rewritten as, $R_{H3}(h) = \Sigma_{\tilde{T} \in \tilde{T}}w_{\tilde{T}}\Sigma_{t \in \tilde{T}}h_{t-1,t}$.

A criticism of this as a measure of reliability is that it assumes the weights $w_t$ are constant, whereas they are likely to decrease as headway increases. To maximize this weighted sum, give as much headway as possible to trains having the largest $w_t$, then give as much as possible to trains having next largest $w_t$, and so on. Trains with small weights $w_t$ may get no headway.
Instead of using the weighted mean headway as a measure of reliability we can use a different average, for example the weighted median headway, or weighted modal headway. Also, recall that \( h_{tt'} = (h_{rt'} - h_{rt'}) \), where \( h_{rt'} \) is the scheduled headway, \( h_{rt'} \) is the minimum required headway, and \( h_{tt'} \) is the residual free headway, between train \( t' \) and the preceding train \( t \). Using this, the reliability measure can be rewritten as,

\[
R_{H3}(h) = \sum_{t=1}^{T} w_{rt'} (h_{rt'} - h_{rt'}) = \left( \sum_{t=1}^{T} w_{rt'} h_{rt'} \right) - \text{(a constant)}.
\]

### 3.4. Heuristics based on line conflicts

When a train is arriving at, or departing from, a platform it is quite common for its scheduled path to cross lines or paths that are used by trains going to or from other platforms. Unscheduled delays to these other trains can delay train \( t \). The more paths train \( t \) has to cross the more likely it is to be delayed, hence we can define a measure of unavailability for train \( t \) as, \( ur_{H4} = (\text{number of train paths crossed by train } t) \). This is a number \( \geq 0 \), hence to convert it to a measure of reliability between 0 and 1 we can write,

\[
r_{H4,t} = 1/[1 + \text{number of train paths crossed by train } t].
\]

This equals 1 if no train paths are crossed by train \( t \) and tends to 0 if a large number of paths are crossed by train \( t \).

To make this measure more sensitive, we can introduce a weight for each of the lines crossed by train \( t \)—giving more weight to lines which are more likely to cause conflicts with train \( t \). Then the above measure of reliability becomes,

\[
r_{H4,t} = 1/[1 + \text{weighted number of train paths crossed by train } t].
\]

For example, let the weight for a line be the number of trains scheduled to use the line near to the scheduled time of train \( t \) (from say 30 min before to 10 min after this time). Or use some other indicator of line activity, for example the fraction of this interval that the line is in use, including the required minimum headway times of trains on the line.

To obtain an aggregate measure of reliability for the set of trains \( T \), take the mean of above individual train reliabilities \( r_{H4,t} \), thus,

\[
R_{H4}(h) = \left( \frac{\sum_{t=1}^{T} r_{H4,t}}{T} \right) / T = \sum_{t=1}^{T} w_{rt} r_{H4,t}.
\]

### 4. Concluding remarks

Most of the measures of reliability in Sections 2 and 3 are applicable to various modes of scheduled transport, but are most relevant in modes where knock-on delays are an important cause of unpunctuality or unreliability. Some of the best, and most complex, examples of knock-
on delays are for trains, hence we use trains as examples. Most of the measures (whether or not using probabilities) involve headways, and the basic reason for this is that longer headways generally reduce knock-on delays.

Knock-on delays may be associated with operating rules or policies, or infrastructure layout. Operating rules or policies include requiring minimum headways between vehicles, or requiring that the next service uses the same driver or vehicle. Infrastructure layout here includes facilities which cannot be used simultaneously by two services (for example train tracks, station platforms, runways, boarding facilities), so that delays to one service may delay the next.

An exact detailed probabilistic analysis of knock-on delays could require a large amount of data, including details of the dispatching rules or policies, infrastructure layout, etc., causing the knock-on delays. For example, if service \( i \) is scheduled to depart before service \( j \) and service \( i \) is late then the rule may be that service \( j \) should wait no more than say 5 min before departing. Since there are such a variety of rules, etc., we do not propose to explore all of them in this paper, especially since we are concerned here with heuristic shortcut measures rather than exact measures. Hence in the probabilistic sections (2.2–2.5) we consider only some typical simple examples. Some further probabilistic analysis of knock-on delays is given in Carey (1994) and Carey and Kwicinski (1995).

We plan to further test the heuristic measures of reliability developed in this paper, by comparing their results with the results obtained from using detailed stochastic simulation models. The latter give accurate measures of reliability, if we make the simulation model sufficiently detailed and run it sufficiently long. For heuristic measures to be useful they do not have to give the same numerical values as simulation based measures. They merely have to give results which are strongly correlated with, or monotonically increasing with, accurate simulation based measures, when both are applied to the same set of schedules. To test the heuristic measures we plan to apply them, and the simulation based measures, to the same set of schedules. If both approaches rank the schedules in the same reliability order, this shows that the heuristic measures are useful consistent measures of reliability. It does not prove they will always be so, but it gives more confidence in using them. The advantage of using heuristic measures is of course, as noted in the introduction, that they can be computed more easily than detailed simulations, and they require less data, which may not be available.

As well as using the heuristic measures developed in this paper to measure the reliability of existing or proposed schedules we can use some of them to assist in the process of generating more reliable schedules. We propose to incorporate some of the heuristic measures of reliability into the scheduling process or scheduling algorithms. Then, when comparing and evaluating options and choices in the scheduling process, we can use local heuristic measures of reliability to guide the choices, in conjunction with using other measures of cost or benefit.

Acknowledgements

This research was supported by UK Engineering and Physical Sciences Research Council (EPSRC) grants GR/H/48033 and G11/K/75798. The author also wishes to thank various British Rail and Railtrack personnel for their advice and cooperation, and the referees for their helpful comments.
Appendix 1: Properties of heuristic measure of reliability $R_{h_{12}}^{w}(h)$ used in Section 2

When train planners or operators are setting train headways it is of interest to know what distribution of headways will yield the most reliable schedule. Here we derive the headway distribution $\hat{h} = [h_{1,2}, h_{2,3}, \ldots, h_{T-1,T}]$ which yields the maximum schedule reliability, as measured by $R_{h_{1}}(h)$ from (4) or $R_{h_{1}}^{w}(h)$ from (17) in Section 2. Recall that,

$$
R_{h_{1}}^{w}(h) = \left( \sum_{t=1}^{T} w_{t} r_{h_{t},t}(h_{t'}) \right) / T = \left( 1 / T \right) \int_{-\infty}^{m_{t}^{*} + h_{t'}} \omega f_{t}(\tau) \, d\tau.
$$

where all weights $w_{t}$ are nonnegative.

**Proposition A1.** Let $m_{t}$ and $m_{t}^{*}$ be the actual and scheduled movement times, respectively, of train $t$, let $f_{t}(m_{t})$ be unimodal and downward sloping at $m_{t}^{*}$, i.e. $f_{t}'(m_{t}^{*}) < 0$. Then $R_{h_{1}}^{w}(h)$ is strictly concave. If $f_{t}'(m_{t}^{*}) \leq 0$ then $R_{h_{1}}^{w}(h)$ is concave.

**Proof.** The second derivative of $R_{h_{1}}^{w}(h)$ w.r.t. $h_{t'}$ is $w_{t} f_{t}'(m_{t}^{*} + h_{t'})$ for all $t'$. If $f_{t}(m_{t})$ is unimodal and $f_{t}'(m_{t}) < 0$ at $m_{t}^{*}$ then $f_{t}'(m_{t}) < 0$ for all $m_{t} \geq m_{t}^{*}$, hence $f_{t}'(m_{t}^{*} + h_{t'}) < 0$ for all $h_{t'} > 0$, hence $R_{h_{1}}^{w}(h)$ is strictly concave. Similarly, if $f_{t}'(m_{t}^{*}) \leq 0$ then $f_{t}'(m_{t}^{*} + h_{t'}) \leq 0$ and $R_{h_{1}}^{w}(h)$ is concave.

**Remarks on assumptions.** It is assumed in Propositions A1 and A2(i)(a) that the movement time pdfs $f_{t}(m_{t})$ are unimodal. This is usually assumed elsewhere and is commonly true in practice. It is also assumed in Propositions A1 and A2(i)(a) that the movement time pdfs are not upward sloping at the scheduled movement time, i.e. $f_{t}'(m_{t}^{*}) \leq 0$. In practice the distribution $f_{t}(m_{t})$ usually has a much longer right tail than left tail (as in Fig. 1). Hence if the scheduled time $m_{t}^{*}$ were in the upward sloping part of $f_{t}(m_{t})$ most of $f_{t}(m_{t})$ would be to the right of $m_{t}^{*}$, so that the actual movement time $m_{t}$ would almost always be late (later than scheduled). Transport planners and schedulers normally set scheduled times so as to avoid such persistent lateness, hence they are implicitly setting $m_{t}^{*}$ to the right of the right tail of $f_{t}(m_{t})$.

![Fig. 1. Illustrative pdf for actual arrival time $m_{t}$, showing scheduled arrival time $m_{t}^{s}$ and headway $h_{tt'}$.](image-url)
right of the peak of \( f_t(m_t) \), even if they are quite unaware of the concept of probability distributions. A typical pdf \( f_t(m_t) \) is shown in Fig. 1.

Proposition A2. Let \( m_t \) and \( m_t' \) be the actual and scheduled movement times respectively of train \( t \), and \( h_{tt'} \), be the headway between \( t \) and the next train movement \( t' \). The sum of the free headways between the series of train movements \( t = 1, \ldots, T \) is fixed at \( H \), i.e., \( \sum_{t=1}^{T} h_{tt'} = H \).

Remark on the latter assumption. Suppose that the number of train movements \( (T) \) is fixed, and the minimum required headways \( (h_{tt'}) \) between train movements are fixed, and the total time spanned by the set of trains (for example, from 8.00 to 10.00 a.m.) is fixed—call it \( H \). This implies that the remaining free time between trains is fixed (i.e. \( \sum_{t=1}^{T} h_{tt'} = H - \sum_{t=1}^{T} h_{tt'} \)).

(i) Let the movement time pdf \( f_t(.) \) be unimodal and \( f_t(.) \leq 0 \) at \( m_t \). Then:

(a) At a maximum of \( R^w_h(h) \), \( w_t f_t(m_t + h_{tt'}) \) is the same for every train \( t' \). If the weights \( w_t \) are equal and the pdf \( f_t(.) \) is the same for some or all trains \( t' \), i.e. \( f_t(m_t' + \hat{\tau}) = f(m_t' + \tau) \), then \( R^w_h(h) \) is maximum when headways \( h_{tt'} \) are equal for these trains.

(b) If \( f_t(.) < 0 \) at \( m_t' \), any maximum is a unique global maximum, and if \( f_t(m_t') \leq 0 \) any maximum is a global maximum but may not be unique [there may be a convex set of \( h \) any of which maximise \( R^w_h(h) \)].

(ii) Let the movement time pdf \( f_t(.) \) for some trains \( t \) be upward sloping at the scheduled movement time (i.e. \( f_t(m_t') > 0 \)). Then:

There may be a local maximum of \( R^w_h(h) \) at \( h_{tt'} = 0 \) for train \( t' \). If the pdf is also unimodal for train \( t \) then there are at most two local maxima for train \( t' \), one at \( h_{tt'} = 0 \) and one as in part (i). Part (i) still holds for the remaining trains.

(iii) Let the movement time pdfs be uniform, i.e. \( f_t(m_t') = \) a constant \( K_t \), over the relevant range \( m_t' \leq m_t \leq U_t \), where \( U_t \) is an upper time limit, and \( f_t(m_t') = 0 \) thereafter. Then:

(a) \( R^w_h(h) \) reduces to \( R^w_h(h) = \sum_{t=1}^{T} w_t K_t h_{tt'} \) where \( 0 \leq h_{tt'} \leq (U_t - m_t') \).

(b) To maximize \( R^w_h(h) \) proceed as follows. Rank the trains in descending order of the constants \( w_t K_t \). Allocate headway to the first train up to the upper limit of its pdf range (i.e. set headway \( h_{tt'} = (U_t - m_t') \), similarly for second train, then the third, and so on until all headway \( H \) has been used up, or until all trains have been assigned headway. If there is any headway left over, it can be divided arbitrarily between trains.

When ranking the trains in descending order of \( w_t K_t \), if some or all trains have the same \( w_t K_t \), rank them in arbitrary order. If all trains have the same pdf value \( w_t K_t \) and if there is not enough headway \( H \) available to give every train its maximum desired headway \( [h_{tt'} = (U_t - m_t')] \), then how this headway \( H \) is divided among trains has no effect on \( R^w_h(h) \)—the allocation of headway between trains is arbitrary.

Proof. (i)(a) To find the maximum of \( R^w_h(h) = \sum_{t=1}^{T} w_t r_{h,t'}(h_{tt'}) \) subject to \( \sum_{t=1}^{T} h_{tt'} = H \), consider the Lagrangian \( L(h) = \sum_{t=1}^{T} w_t r_{h,t'}(h_{tt'}) + \lambda (\sum_{t=1}^{T} h_{tt'} - H) \). The assumptions in (i) ensure \( R^w_h(h) \) is concave (from Proposition A1). Hence the maximum occurs when \( \partial L(h)/\partial h = 0 \), i.e. \( w_t f_t(m_t' + h_{tt'}) - \lambda = 0 \) for all \( t' \), hence \( w_t f_t(m_t' + h_{tt'}) = \lambda \) for all \( t' \), hence \( w_t f_t(m_t' + h_{tt'}) t \) is the same for all \( t' \).

If also \( f_t(m_t' + \tau) = f(m_t' + \tau) \), then \( w_t f_t(m_t' + h_{tt'}) \) is the same for all \( t' \) when all headways \( h_{tt'} \) are equal, hence \( R^w_h(h) \) is maximum when all headways are equal and all \( w_t \) are equal.
(b) If \( f'_t(m_t) < 0 \) then, from Proposition A1, \( R^w_h(h) \) is strictly concave hence any maximum is a unique global maximum, and if \( f'_t(m_t) \leq 0 \), then \( R^w_h(h) \) is concave hence any maximum is a global maximum but may not be unique.

(ii) If the movement time pdf \( f_t(m_t) \) is upward sloping at the scheduled movement time \( m_t^* \), then (see proof of Proposition A1) \( R^w_h(h) \) is convex from there until \( f_t(m_t) \) stops increasing or begins to slope down. The maximum of any convex function is at a corner point, hence if \( R^w_h(h) \) is initially convex it may have a local maximum at a corner point, i.e. at \( h^*_t = 0 \) for some trains \( t' \).

(iii) The results (a) and (b) follow immediately from the definition of \( R^w_h(h) \). \qed

Appendix 2: Some properties of \( R_{H2,\hat{T}} \) in (22)–(25)

In Proposition 2 in Section 3.2 we considered the effect on the reliability measures (22)–(23) of changes in the aggregate headway \( H_{\hat{T}} \) for trains of type \( \hat{T} \). Here we consider the effect on (22)–(23) of changes in the number \( n \) of trains of type \( \hat{T} \). We wish to check that this does not change the reliability measure in the “wrong” direction. Changing the number of trains changes the mean headway per train, hence to make a fair comparison of schedules with different numbers of trains, we assume they have the same mean headway and same headway m.a.d. Then:

**Proposition A3.** Suppose we

(a) multiply the number of trains by \( k \) and multiply the total headway by \( k \) so that the mean headway remains the same, and

(b) assign the additional trains and headway so that the headway m.a.d. remains the same.

Then the numerical values of unreliability \( (1 - R_{H2,\hat{T}}) \) from (23) is multiplied by a factor \((n-1)/(n-(1/k))\). If \( n \) is large relative to \((1/k)\) this factor is approximately 1.

**Proof.** By Assumption (b), the numerator of \((1 - R_{H2,\hat{T}})\) (i.e. m.a.d.) is unchanged. Substitute \( kn \) and \( kH_{\hat{T}} \) for \( n \) and \( H_{\hat{T}} \), respectively, in the denominator of \((1 - R_{H2,\hat{T}})\), compare the new and old values of \((1 - R_{H2,\hat{T}})\) and cancel the \( k \). \qed

**Proposition A4.** As Proposition A3 except in (b) assign headways so as to leave s.d. unchanged rather than m.a.d. unchanged.

Then the numerical value of unreliability \( (1 - R_{H2,\hat{T}}) \) from (22) is multiplied by a factor \( \sqrt{(n-1)/(kn-1)} \). If \( n \) and \( nk \) are large this ratio is approximately \( 1/k \).

**Proof.** As for Proposition A3. \qed

We can revise the reliability measures (22) and (23) so that the changes (a) and (b) in Propositions A3 and A4 will have no effect on the value of reliability. This gives (24) and (25). Hence:

**Proposition A5.** Assume (a) and (b) from Proposition A3.

Then the numerical values of reliability \( R_{H2,\hat{T}} \) from (24) and (25) are unchanged.

**Proof.** As for Proposition A3.
However, the maximum values of reliability in (24) and (25) are \((n - 1)/n\) and \(\sqrt{n - 1}\), respectively, rather than 1. This should be borne in mind when comparing reliability of schedules having different numbers of trains \(n\). Unless the number of trains \(n\) is very small, the maximum values of reliability in (25) [i.e. \((n - 1)/n\)] is approximately 1 and for that reason (25) is a more natural measure of reliability than (24).

References


