UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2013:6

Stability conditions for scheduled waiting time in railway traffic

Rikard Molander

Examensarbete i matematik, 15 hp
Handledare och examinator: Ingemar Kaj

April 2013

Department of Mathematics
Uppsala University

Stability conditions for scheduled waiting time in railway traffic

Rikard Molander

April 21, 2013

1 Introduction

Train traffic, following a set time-table, may be considered more predictable than other forms of traffic. However, the requests for usage of the railway system arrive no more predictably than do cars on a highway or data packets in a computer network. Therefore, the application of queuing theory to time-table design may yield useful results in predicting so-called *scheduled waiting time*, that is, the difference in time between a trains' intended departure time and the first clear space in the time-table. (Wendler [3])

In this paper, we will consider the expected values and the stability conditions for scheduled waiting time on one and two tracks, respectively, closely following the method laid out for aircraft by Bäuerle et. al. We will regard the headway times - the minimum time needed between two departures - as service times in a queuing system.

2 Terminology

A *traffic node* is a point of interest on a railway track. It may be a train station, a place where two tracks merge, or simply an arbitrary point on a railway track. For purpose of this paper, we make two assumptions about traffic nodes;

1. Traffic on the node is one-way. We consider only the option of passing the node in a single direction.

2. Leaving the node is possible via either one or two tracks. In section 4 we consider a single track; in section 5, a double track.

By *headway*, we mean the smallest safe distance between one train and another. By *headway time*, we mean the time it takes after train A passes a certain point before train B can safely follow it. Obviously, the headway time depends directly on the headway. Both headway and headway time depend on a number of factors, the most dominant being the relative speeds of the train and their braking distance. For instance, a fast-moving, heavily loaded train might need a long stretch of headway before it, while a slow-moving, light-loaded train might need very little.

3 Basic model

In the most simple case, assume that we have a single traffic node with a single track that can be used by one train at a time. The train departs from this node, and due to safety regulations a certain time must pass before another train can safely follow it. This elapsed time is referred to as the *headway time*. The node is considered "occupied" for the duration of the headway time, and if any new trains arrive during this period, they must wait, forming a queue. Of course, this is not a physical queue as the process of scheduling train departures does not occur in real time - but as train-operating companies put in requests to use the tracks, scheduled departures are moved forward in time to accommodate other, earlier trains, creating a classical queueing process.

In the most basic case, we can consider the process to be M/M/1. Here, we consider the incoming trains as customers, arriving in a Markovian fashion, while the minimum headway times are service times (the time it takes for a traffic node to "process" a train). The service times are randomly distributed, as each train varies in speed, load, state of repair et cetera, and they are also considered independent from each other.

If this is the case, the stability condition is easy to find. If we denote the arrival intensity by λ and let the service times be $\sim Exp(\mu)$, then the system is stable iff $\rho := \frac{\lambda}{\mu} < 1$. The expected waiting time for a customer (i.e. an arriving train) is determined by Little's law to $W := E(W_t) = \frac{1}{\mu - \lambda}$.

This model is, however, overly simplistic. Specifically, we cannot assume the service times to be independent, as they represent the minimum time needed between one train and the next. A slower train preceding a faster train will necessitate a longer headway time, and vice versa, a faster train preceding a slower one will necessitate a much shorter headway time. Other factors, like differences in braking distance, may also play into the difference in headway time. Therefore, a more advanced model must be used, forming the bulk of this paper.

4 Headway times for a single track

We shall here draw on a model used by Bäuerle et. al. [2] for air plane runways.

Trains arrive according to some arrival process at time S_n and with inter-arrival times $T_n = S_n - S_{n-1}$, with $S_0 = T_0 = 0$, and the first train arriving at time $S_1 = T_1$.

In reality, there may be many factors playing into the headway time required between two trains, but for our model it is sufficient to assume that we have a discrete amount of differing train types; denote this amount k. The probability that a given train is of type i is denoted p_i, and of course, $\sum_{i=1}^{k} p_i = 1$. We assume that trains arrive according to a Markovian arrival process with intensity λ where the train types are independent of each other, so that if J_n denotes the type of the nth train, then $P(J_n = i | J_{n-1} = j) = P(J_n = i) = p_i$.

If the arrival process is a Poisson process, then we may view the arrival process as a superposition of k independent Poisson processes each with intensity $\lambda_i = \lambda p_i$.

As a train of type i rolls away from the traffic node, a certain headway time must pass before a train of type j may follow it. Denote this headway time as $b(i, j)$, so that at least $b(i, j)$ time units must pass between the departure of the first train and the departure of the second. This gives rise to a $k \times k$ matrix of headway times, denoted C, which we shall consider to be given. In practice, such a matrix can be obtained by measuring e.g. the typical speed and braking distance for the various kinds of trains.

4.1 Service times

As mentioned, the service time in our model is the headway time between two trains. Therefore the service time of the nth train is the headway time between it and the following train $n + 1$, so that the service time can be denoted $B_n := b(J_n, J_{n+1})$, with $B_0 := 0$. When a request for a train to use the traffic node arrives, the headway time may have expired - in which case there is no waiting time - or the new train may have to wait a certain period, denoted W_n for the nth train. The service times B_n are randomly, and identically, distributed - but they are not independent, since B_{n+1} by definition depends on B_n.

4.2 Stability

When the (n+1)th request to use the traffic node arrives, it may be displaced in time by some waiting time W_n. This time period is determined recursively by

$$W_1 := 0; W_{n+1} := [W_n + B_n - T_n]^+$$

where $n \geq 1$. To find whether or not the queuing system is stable, we need to determine whether the sequence (W_n) is stable.

By Loynes [1] Corollary 1, we know that if the sequences (B_n) and (T_n) are independent of each other, and if one of them is formed of non-constant mutually independent random variables, then the sequence (W_n) is stable if $E(B_1) < E(T_1)$, with $n \geq 1$, and unstable otherwise.

In our case, we have that (T_n) is formed of non-constant mutually independent random variables, as the arrival process is assumed to be Markovian. Hence, the condition holds.

We may denote the condition by re-introducing the variable ρ, such that

$$\rho := \frac{E(B_1)}{E(T_1)}$$

and conclude that the queue is stable iff $\rho < 1$.

It then remains to compute ρ.

The value of $E(T_1)$ is known; the arrival process is assumed to be a Poisson process with intensity λ, so that $E(T_1) = \frac{1}{\lambda}$.

The value of $E(B_1)$ is somewhat more complicated. B is given by the function $b(J_1, J_2)$, which is in itself a random variable, with $E[b(J_1, J_2)] = \sum_i^k \sum_j^k p_i p_j b(i,j)$. As all $b(i,j)$ are represented in the matrix C, we can easily form a new matrix D of the same dimensions, with entries $d(i,j) := p_i p_j b(i,j)$. Then

$$E(B_1) = E[b(J_1, J_2)] = \mathbf{1}^T D \mathbf{1}$$

where $\mathbf{1}$ denotes the vector of dimension k containing solely ones, and hence that

$$\rho := \frac{E(B_1)}{E(T_1)} = \lambda \mathbf{1}^T D \mathbf{1}$$

Therefore, if $\lambda \mathbf{1}^T D \mathbf{1} < 1$, then the queueing system is stable.

Using the above result, we may calculate the highest possible intensity of the arrival process for the queueing system to remain stable. If $\lambda < 1/\mathbf{1}^T D \mathbf{1}$, then the system remains stable.

5 Model for double tracks

We now turn to the case where there are two tracks succeeding the traffic node. In this case, we may be able to handle a higher traffic intensity. With two tracks, an incoming train J_i may utilize either one of two tracks, denoted I and II. We assume that these two tracks cover the entire distance considered, so that the tracks never need to merge - if so, that could cause scheduling conflicts further down the line, which is undesirable.

The arrival process remains unchanged, so that T_n is the same as in the case with one track. Only the service time B_n is altered by the presence of two tracks. As it turns out, however, this model rapidly becomes quite complex: If the first train utilizes track I, and the second train utilizes track II, then the model is unusable because B_1 will have no impact on the waiting time for the second train. Hence our previous method will no longer be applicable.

There are a few ways we can deal with this problem.

5.1 Why not pick the free track?

A natural question may be, why not pick the track with the shortest waiting time upon arrival? There chief reason why this approach isn't considered is mathematical; it is very difficult to design such a model, since service times are not independent of each other, and it is therefore difficult to predict the waiting times. The method presented by Loynes [1] does not hold under these circumstances, and it is unknown what type the next train will be, further complicating this approach.

5.2 Alternating Tracks

The simplest, and most straightforward solution, is to simply alternate which track is used. Bäuerle et. al refer to this as the *Round-robin policy* (Bäuerle et. al. [2]). Odd-numbered trains are assigned to track I, while even-numbered trains use track II.

We then have two sequences $J_m^I = J_{2m-1}$ and $J_m^{II} = J_2 m$, with J_n denoting the type of train, each with separate arrival intensities and separate average waiting times. Notably, this has no impact on the headway times; therefore, it still holds that $E(B_1^I) = E(B_1^{II}) = \mathbf{1}^T D\mathbf{1}$. Hence, we can conclude that the separate processes J_m^I and J_m^{II} are such that, with $a \in I, II$

$$\rho^a := \frac{E(B_1^a)}{E(T_1^a)} = \lambda^a \mathbf{1}^T D\mathbf{1}$$

Therefore, track I is stable iff $\lambda^I < \frac{1}{\mathbf{1}^T D \mathbf{1}}$, and likewise for track II, and therefore the supremum of the intensity for track I is $\frac{1}{\mathbf{1}^T D \mathbf{1}}$ and likewise for track II. Since the arrival processes for both tracks are Poisson, we can merge them together and create an arrival process for the whole system. As $\lambda^I = \lambda^{II}$, it is the case that this greater system will have the arrival intensity $\lambda = \lambda^I + \lambda^{II} = 2\lambda^I = \frac{2}{\mathbf{1}^T D \mathbf{1}}$.

As expected, with two tracks and trains assigned in an alternating fashion, the highest possible intensity is twice as high as in the single-track case. This should hardly be surprising - it simply says that two tracks can handle twice as many trains as one.

5.3 Random Assignment

It is also possible to randomly assign incoming trains to a track. We could do this either perfectly randomly - by giving each incoming train a .5 chance of choosing track I, say - or we could sort the trains by type, so as to let e.g. heavy trains favour track I and light trains favour track II, or some similar method. The former case is relatively uninteresting, as it does not differ much from the alternating policy (Bäuerle et. al. [2]), but we can inspect the latter a little more thoroughly.

Let $\delta_j \in [0, 1]$ denote the probability that a train of type j is assigned to track I, and consequently, $1 - \delta_j$ is the probability to be assigned to track II. Because the arrival process is a Poisson process, it can be split using this policy. Let us study the case for track I.

For track I, each incoming train is of type j with probability p_j; furthermore, a train of type j is assigned to track I with probability δ_j, as described above. There are k different types of trains. Then the arrival intensity for track I can be defined by

$$\lambda^I := \lambda \sum_{j=1}^{k} p_j \delta_j$$

(Note that, if δ_j is 1 for all j, we have the case for a single track, where trivially $\lambda = \lambda \sum_{j=1}^{k} p_j$ because $\sum_{j=1}^{k} p_j = 1$ by definition.)

We can express this more conveniently in vector form, by writing $p = (p_j)$ and $\delta = (\delta_j)$, each forming a vector with k elements. Then $\lambda^I = \lambda p^T \delta$ and conversely, $\lambda^{II} = \lambda(1 - p^T \delta)$. We've now formed two separate, independent Poisson processes, each operating on its own track and behaving as in the single-track case.

5.3.1 Stability on two tracks

Recall that, by Loynes [1], the sequence is stable if $E(B_1) < E(T_1)$ or equivalently $\rho := \frac{E(B_1)}{E(T_1)} < 1$. The first problem, then, is to find $E(T_1)$ and $E(B_1)$ for each of the two new processes.

$E(T_1)$ is simple; this is given for each individual track by $\lambda^{I(-1)}$ and $\lambda^{II(-1)}$, respectively. Now let us attempt to find $E(B_1)$.

The probability that a certain train arriving at track I is type j is given by $p^I_j := \frac{p_j \delta_j}{p^T \delta}$ (because of the laws of conditional probability, $P(A|B) = \frac{P(A \cap B)}{P(B)}$), and similarly for track II, $p^I I_j := \frac{p_j(1-\delta_i)}{p^T(\mathbf{1}-\delta)}$. From this we can compute the expected service times, using a similar method as in the one-track case.

In the case of a single track, we had $E(B_1) = E[b(J_1, J_2)] = \sum_i^k \sum_j^k p_i p_j b(i,j)$. Now, looking at Track I, we must replace p_i and p_j with p^I_i and p^I_j, so that $E(B^I_1) = \sum_i^k \sum_j^k p^I_i p^I_j b(i,j)$ Now, the values $p_i p_j b(i,j)$ are given by the matrix D, so that this can be expressed as

$$E(B^I_1) = \sum_i^k \sum_j^k p^I_i p^I_j b(i,j) = \frac{\delta^T D \delta}{(p^T \delta)^2}$$

Which is similar to the case with a single track, except that the vectors $\mathbf{1}$ have been replaced with the vectors $\frac{\delta}{p^T \delta}$. Track II is handled analogously, with $E(B^{II}_1) = \frac{(\mathbf{1}-\delta)^T D(\mathbf{1}-\delta)}{(1-p^T\delta)^2}$.

Putting this together, we have that

$$\rho^I := \frac{E(B^I_1)}{E(T^I_1)} = \frac{\delta^T D \delta}{(p^T \delta)^2} / (\lambda p^T \delta)^{-1} = \lambda \frac{\delta^T D \delta}{p^T \delta}$$

and

$$\rho^{II} = \frac{E(B^{II}_1)}{E(T^{II}_1)} = \lambda \frac{(\mathbf{1}-\delta)^T D(\mathbf{1}-\delta)}{1 - p^T \delta}$$

The system, then, is stable iff both ρ^1 and ρ^{II} are both less than 1, that is, iff $\max\left\{\rho^I, \rho^{II}\right\} < 1$. Equivalently, the supremum of the arrival rate λ_{sup} is then given by:

$$\lambda_{sup} = min\left\{\frac{p^T\delta}{\delta^T D \delta}, \frac{1 - p^T \delta}{(\mathbf{1}-\delta)^T D(\mathbf{1}-\delta)}\right\}$$

If the arrival rate is smaller than this value λ_{sup}, then the system is stable.

6 A Simple Example

Consider a system with three kinds of trains, type A, B, and C. Type A trains are heavily-loaded cargo trains, type B are passenger trains, and type C are express trains. The matrix C detailing their respective headway times is such:

$$\begin{pmatrix} 8 & 16 & 24 \\ 4 & 8 & 12 \\ 2 & 4 & 6 \end{pmatrix}$$

With values listed in minutes. Furthermore, the traffic is such that $p_A = 0.3$, $p_B = 0.5$ and $p_C = 0.2$.

From this, we can compose the matrix D, combining the probabilities and the headway times:

$$\begin{pmatrix} 0.72 & 2.40 & 1.44 \\ 0.60 & 2.00 & 1.20 \\ 0.12 & 0.40 & 0.24 \end{pmatrix}$$

In the case of a single track, we can quite easily compute the highest possible arrival intensity. The traffic intensity is given by $\frac{E(B_1)}{E(T_1)} = \lambda \mathbf{1}^T D \mathbf{1}$ and conversely, then, the arrival intensity must be $1/(\mathbf{1}^T D \mathbf{1})$. This is readily computed using matrix multiplication; the value is 1/9.12, that is, an average of roughly 0.1096 trains per minute, or 6.579 trains per hour. Therefore, under these circumstances, one train every 10 minutes results in a stable system.

If we were to use two tracks and simply alternate them, we could run twice as many trains - roughly one every five minutes, or 13.158 trains per hour.

Now let us construct a simple semi-random double-track strategy. In this variant, type A trains always choose track I, and type C trains always choose track II, whereas type B trains are routed to track I 32.8% of the time. This means that the vector δ, representing the probability of a train using track I, contains the elements (1, 0.328, 0). The vector p is given above as $[0.3, 0.5, 0.2]$.

We calculate λ_{sup} by using the formula outlined above:

$$\lambda_{sup} = min \left\{ \frac{p^T \delta}{\delta^T D \delta}, \frac{1 - p^T \delta}{(1 - \delta)^T D (1 - \delta)} \right\}$$

In this case,

$$\frac{p^T\delta}{\delta^T D\delta} = \frac{0.4640}{1.919} \approx 0.2418$$

and

$$\frac{1 - p^T\delta}{(1-\delta)^T D(1-\delta)} = \frac{0.5360}{2.2184} \approx 0.2416$$

The minimum of this is 0.2416, that is, 0.2416 trains per minute, or 14.49 trains per hour. This is a slight improvement over the alternating track method, showing that a more careful selection process may be beneficial.

7 Summary

In the study of train traffic, the headway times between trains depend on the different types of trains used. A preceding train might require more or less time to clear a safe distance, requiring a queueing model in which service times cannot be considered independent of each other; to solve this problem, a method similar to that used by Bäuerle et. al. on airway traffic was described. This method is straightforward to apply to train traffic in the case of a single track; it is also applicable, with some modification, to a double-track situation.

Various strategies can be used to route incoming train traffic between the two tracks. An alternating model was used, as well as a probabilistic model where incoming trains are assigned according to some previously determined probabilities. Finally, the model was applied to a concrete example concerning three types of trains.

References

[1] R.M Loynes,The stability of a queue with non-independent arrival and service times, Proceedings of the Cambridge Philosophical Society 58 (1962) 497-520.

[2] N. Bäuerle, O. Engelhardt-Funke, M. Kolonko, On the waiting time of arriving aircrafts and the capacity of airports with one or two runways, European Journal of Operational Research 177 (2007) 1180-1196.

[3] E. Wendler, The scheduled waiting time on railway lines, Transportation Research Part B 41 (2007) 148-158.