Increasing Robustness of Dense Timetables by Visualization of Train Traffic Record Data and Monte Carlo Simulation

Kohei USHIDA
Tokyo Metro Co., Ltd., 3-19-6 Higashi-Ueno, Taito-ku, Tokyo 110-8614 Japan

Shingo MAKINO, Norio TOMII
Chiba Institute of Technology, 2-17-1 TsudanumaNarashino-shi, Chiba 275-0016 Japan

Abstract: We propose an algorithm to obtain a more robust timetable by finely tuning a given timetable. Firstly, we propose methods to visualize train traffic record data. Using the methods such as a Chromatic diagram, we can figure out where delays are occurring and how they are propagating at a glance. Secondly, we propose an index which we call a Buffer Index. A Buffer Index is calculated from a delay and a buffer time for each train and station and represents a characteristic of a delay from a viewpoint of causing knock on delays. Buffer Indices are used as a clue to know how to modify a timetable to make it more robust. Finally, we introduce an algorithm to modify a given timetable to make it more robust. The algorithm is based on the Monte Carlo Simulation. As for the simulation algorithm, we introduce an algorithm which we call a fact based simulation. In the fact based simulation, durations between events are decided from analytical results of train traffic record data. We have confirmed that our algorithm is quite promising using real world timetable data.

Keywords: Robustness, Chromatic Diagram, Buffer Index, Monte Carlo Simulation, Timetable

1 Introduction

In urban areas of Japan, railways play quite an important role for commuting. As a matter of fact, the market share of railways in three major cities of Japan is almost one half. In order to transport such a massive amount of passengers, trains are operated quite densely. In many railway lines in big cities, 25 - 30 trains are running per hour per direction on a double track line during morning rush hours. This means trains are running every two to three minutes in one direction. Still, trains are very congested and it is not unusual that more than 2,000 passengers are aboard a train which consists of ten cars and is 200 m long.

In order to operate so many trains with such a short interval, timetables are made very carefully. Timetables are usually made with a unit of five seconds. Timetable planners are very cautious to avoid conflicts imposed by capacity constraints. They must pay attention to dwell times of trains because dwell times must not be too long to pack in a number of trains within a certain time of period, 25 trains per hour, for example. On the other hand, they must be long enough for passengers getting on and off. In many railway lines, dwell times of trains are typically 30 - 40 seconds for intermediate stations. Dwell times are cautiously defined considering the number of passengers who get on/off the trains.

During daytime, it is common to adopt a periodic timetable, but during morning and evening rush hours, a non periodic timetable is introduced in order to meet the passengers’ demand.

Currently, one of the serious problems in Japanese railways is that small delays happen quite often, almost every day during rush hours. Because trains are operated very densely, once a delay (primary delay) happens, the delay is propagated to other trains and a lot of trains are also delayed (secondary delay). Primary delays are quite often caused by passengers. If more passengers than expected get on/off a train, the dwell time becomes longer and the train is delayed.

It is now desired by railway companies to avoid occurrence of primary and secondary delays by tuning up timetables. It is desired to avoid primary delays caused by passengers by appropriately adjusting dwell times and to avoid secondary delays by appropriately adjusting intervals between trains.

Various factors are relevant to the robustness of timetables. In particular, improvement or increase of facilities such as construction of new tracks is quite effective to increase robustness of timetables, but usually prohibitively expensive and sometimes impossible due to limitation of spaces. So, railway
companies are more interested in improving robustness of timetables by slightly modifying them. Possible measures of modification are: to change departure/arrival times of trains, to modify dwell times, to change tracks, to change stop/pass and so on.

In this paper, we propose an algorithm to obtain a more robust timetable by finely tuning a given timetable. We do not aim at making a timetable from scratch, because timetables are made elaborately reflecting the passengers’ demands.

One of the authors of this paper (Ushida) had a successful experience where he greatly reduced delays of trains for a railway line called TOZAI line of Tokyo Metro Co. Ltd. (hereafter, referred to TOZAI line) by tuning up the timetable. TOZAI line is a subway line which connects the eastern part of Tokyo, the capitol and the western part of Tokyo. The total length of tracks is about 30 km with 23 stations. The number of passengers a day is about 1.32 million and is known as one of the most congested line in Japan as the population along the line increased rapidly. TOZAI line is somewhat different from ordinary subway lines because several types of express trains are operated and half of the track is on the ground. Until Ushida renewed the timetable, even when no particular accidents happen, delays of more than five minutes were usually happening every weekday and the railway company received a lot of complaints from passengers. But after the timetable was renewed, the delays were greatly reduced and the complaints were also decreased to less than one tenth.

The aim of our research is to generalize and computerize Ushida’s know-how. In particular, we introduce an index which we named the Buffer Index, useful to identify which part of the timetable should be amended and how much the change of the arrival/departure times should be [USHIDA10]. Then, we would like to propose a computer algorithm to produce a more robust timetable by slightly modifying the given timetable focusing on the Buffer Indices.

To this end, we introduce methods to utilize train traffic record data which we can obtain from train traffic control systems. Firstly, we propose several methods to visualize train traffic record data which we named Chromatic Diagram, Delay Diagram, 3D Diagram and so on. Using these diagrams, we can figure out where delays are occurring and how they are propagating to other trains at a glance. Secondly, we propose the Buffer Index mentioned earlier. A Buffer Index is calculated from a delay and a buffer time for each train and station and represents a characteristic of a delay from a viewpoint of causing knock on delays. Buffer Index is used as a clue to know how to modify a timetable to make it more robust. Finally, we introduce an algorithm to modify a given timetable and make it more robust. The algorithm is based on the Monte Carlo Simulation. As for the simulation algorithm, we introduce an algorithm which we call a fact based Simulation. The fact based simulation is based on PERT based simulation but durations between events are decided from analytical results of train traffic record data. We have confirmed that our algorithm is quite promising using real world timetable data.

This paper is organized as follows. In section 2, we explain a process to realize a robust timetable. In section 3, we describe our visualization methods. In section 4, we explain the notion of the Buffer Index and its significance for robustness of timetables. In section 5, we introduce an algorithm to obtain a more robust timetable by tuning up departure and arrival times of a given timetable.

2 Process to realize robust timetable

2.1 Robustness of a timetable

An initial delay is defined as a delay which is brought from outside. For example, it is common in the center of big cities in Japan to operate trains directly between suburbs and downtown. So, trains directly go into the subway line. If a train is delayed in the suburbs, the delay is usually brought to subway lines and vice versa. This is a typical example of initial delay.

A primary delay is a delay which is caused by the train itself. Causes of small primary delays are usually passengers. If the time required for passengers’ getting on/off is longer than the dwell time specified in the timetable, the train is delayed. Sometimes, there are reasons for increasing the dwell times. Passengers rush into closing doors and the doors have to be reopened. Some passengers become sick and it takes some time to rescue them. If the number of habitants around a station increases because of redevelopment around the station, it is quite probable that the number of passengers who use the station
also increases and longer dwell times than prescribed in the timetable are needed.

A secondary delay is a delay which is caused by other trains. A knock on delay is a typical example of secondary delay. Due to capacity constraints no more than two trains can use the same track at the same time, the succeeding train is kept waiting outside the station and arrives with a delay. If there are abundant margins in running times and buffer times in interval times between trains, trains are not delayed even if a big initial delay occurred. But in a railway line where a lot of trains are operated densely, this cannot be expected, because it becomes impossible to set up enough trains within a certain time period and railway companies cannot provide a transportation service which meets passengers’ demand. Large buffer times and margins are not desirable from managerial point of view either, because this means that the investment on facilities is not fully used. Hence, it is very important to give necessary but minimum buffer times and margins.

In this paper, we call a timetable robust if occurrence of primary delays is as rare as possible and the propagation of the delay is as small as possible.

Robustness of a timetable has been paid a lot of attention both in Europe and Japan [CAREY99][LIEB07][KROON08][TAKEUCHI06]. In Japan, in urban areas, trains are operated quite densely but still very congested (congestion rate is 150% - 180%). Hence, delays caused by an excess of dwell times often happen and the delays easily propagate to other trains. Hence, railway companies have considerable concern for robust timetables.

2.2 Processes to improve robustness of timetables

We think processes to improve robustness of a given timetable should be as illustrated in Figure 1.

In the Overview phase, we overview if some problems are occurring or not. For example, we overview the average delays and the number of trains that suffer from knock on delays. We observe not only for one particular day, but we are interested in effectively grasping occurrence of problems over a longer term.

In the Identify phase, we identify the cause and the severity of the problem which was found in the overview phase. For example, if Train X is observed to be usually delayed in the overview phase, we identify the cause is Train Y’s delay which is a preceding train of Train X. Typically, we identify an initial or a primary delay as a cause of secondary delay. In urban areas, timetables are quite complicated and various factors are mutually related. Thus, an effective method to overcome the complexity is desirable. In addition, we need to establish a method to quantify the severity of the problem that is used in the Propose phase.

In the Propose phase, we pick up problems based on the severity measure calculated in the Identify phase and propose how we should modify the timetable. To name a few, potential adjustments are:

- Dwell time of Train Z at Station A should be increased by five seconds.
- Running time of Train W from Station B to Station C should be increased by 10 seconds.
- Departing order of Train S and Train T should be changed so that Train T departs first.
- Train U should stop at Station D.

In the Predict phase, we predict results of the modifications taken in the Propose phase. This should be done considering the occurrence of initial and primary delay and propagations of those delays.

In the Evaluate phase, we calculate a quantitative value to evaluate the proposal from the viewpoints of robustness.

In this paper, we introduce methods to visualize train traffic record data, which we believe to be useful in the Overview phase. We propose the Buffer Index as a clue used in the Propose phase. We introduce the Monte Carlo simulation method to assist the Predict phase and the Evaluate phase.
3 Visualizing the Train Traffic Record Data

3.1 Basic Ideas

Train traffic record data are now easily obtained. Train traffic record data contain train numbers, arrival times, departure times at stations and so on, and are collected from signaling equipments. Data for all trains and all stations for a whole day are available over an extended period.

However, it is difficult even for experts to interpret those data because train traffic record data are nothing but lists of numbers. We propose to visualize train traffic record data so that we can overview the underlying problems.

Timetable planners are familiar with so called time space diagrams. Hence, it is a reasonable idea to visualize the train traffic record data in a format of the time space diagram. In addition to ordinary time space diagrams, we propose to use information such as colors, somewhat different formats of time space diagrams, a three dimensional display and other formats which were not used effectively until now.

3.2 Visualization formats

(1) Chromatic Diagram

In ordinary time space diagram, a delay of a train is illustrated as a horizontal difference from its original position. Hence, it is easy to understand that the running time of a train is longer than others from the slant of the train segment for example. But it is not easy to observe how serious delays are and how widely delays are propagating.

The Chromatic Diagram is a two dimensional time space graph similar to the ordinary time space diagrams currently used, but each train segment between station lines is given a color reflecting its delay (departure delay or arrival delay, which is subject to users' choice). We prepare 20 colors from indigo to red (concept of the Chromatic Diagram is illustrated in Figure 2). Chromatic Diagrams make it quite easy for us to visually grasp how serious delays are, where delays are emerging and disappearing and how widely delays are propagating.

(2) Delay Diagram

In ordinary time space diagram, the horizontal axis is the time, whereas in the Delay Diagram, horizontal axis means discrepancy from planned schedule (Concept of the Delay Diagram is illustrated in Figure 3). Thus, if a train is totally on schedule from the start to the end, it is depicted just by one vertical line. From the Delay Diagram, we can easily identify where trains began to be delayed, where delays are recovered and so on. Each train segment is given a color depending on its delay as in the Chromatic Diagram.

(3) Incremental Delay Diagram

In the Delay Diagram, horizontal axis means delays, whereas in Incremental Delay Diagram, the horizontal axis is an increment or a decrement of delays (see Figure 4). So, from the Incremental Delay Diagram, we can learn how delays are growing or decreasing.

![Fig.2: Concept of the Chromatic Diagram.](image)

![Fig.3: Concept of the Delay Diagram.](image)
Challenge F: Even more trains even more on time

(4) 3D Diagram

In the 3D Diagram, horizontal axis (x axis - time) and vertical axis (y axis - station) are the same as in the ordinary time space graph. But 3D Diagram has one more axis (z axis). Z axis depicts delay or dwell time. In other words, a delay or a dwell time of a train at each station is depicted by a pillar. We have also developed a Bubble Diagram with a ball on the pillar in which the ball depicts the dwell time and the pillar depicts the delay. An example of a Bubble Diagram is shown in Figure 5.

3.3 An example of overviewing robustness by visualization methods

We show an example of overviewing robustness of a timetable using the visualization methods. The target line is the TOZAI Line mentioned before.

Before the timetable was renewed in March 2009, TOZAI Line had a serious problem: during morning rush hours, trains were delayed more than five minutes every weekday due to congestion and the delays propagated to a lot of trains. Timetable of TOZAI Line was renewed in order to avoid such delays as follows.

1. Stopping patterns of trains were reconsidered so that all trains stopped at several stations (Kasai - Minami-Sunamachi) which were used by many passengers.
2. Dwell times of trains before morning rush hours were adjusted so that delays of these trains did not influence the trains in the rush hours.
3. Dwell times of regular trains with rather long interval from the preceding stopping train were adjusted (if the preceding train is an express which does not stop, the interval from the second preceding train which stops at the station tend to become longer than usual).
4. Dwell times and intervals between trains
were adjusted by 5 to 10 seconds on the whole, so that there is some amount of buffer.

In addition, the position where trains stop on a track was moved at a station (Monzen Naka-cho) by several meters so that passengers did not concentrate on one door. In some stations, the number of staff on platforms was increased. In some places, signaling systems were upgraded so that the minimum headways between trains became shorter.

In Figures 6 and 7, we show a part of the Chromatic Diagrams for “Before March 2009” and “After March 2009” respectively. In these Chromatic Diagrams, train segments with more than 190 seconds delay are given red color. From these figures, we can intuitively grasp that delays were greatly reduced in “After March 2009” and find that red train segments almost disappeared. Using other visualization methods such as Delay Diagrams, we can analyze that occurrence and propagation of delays were reduced and the excesses of dwell time are less likely to happen than before and so on. Thus, during the process of such analysis, we confirmed that we can effectively use our visualization methods.

4 Buffer Index

In 3.3, we introduced a practical example to make a timetable more robust. In this section, we try to generalize the method used in the process. We propose an index which we call a Buffer Index. A Buffer Index is calculated from the delay of a train and the buffer given to it. We intend to use the Buffer Index as a clue to help us to find which part of a timetable needs to be modified and to know how we should modify it.

4.1 Basic Ideas

Let us assume that a train is delayed for one minute. If the delay does not affect the succeeding trains, the one minute delay does not matter at all. On the other hand, if the succeeding train suffers from a knock on delay due to the one minute delay, the delay cannot be ignored. At the same time, we should consider how often the delay happens, everyday, once a week or seldom.

Severity of a delay is not dependent only on its value but dependent on whether it affects the succeeding trains and how often it happens. Delays we have to urgently cope with are the ones which:

- Bring big knock on delays.
- Happen quite often.

This means that we have to discuss a property of delays as described in Table 1.

<table>
<thead>
<tr>
<th>Knock on Delay</th>
<th>How often</th>
<th>Quite often</th>
<th>Seldom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serious</td>
<td></td>
<td>Urgent Action required</td>
<td>Low Priority</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td>Not a serious problem</td>
<td>No Problem</td>
</tr>
</tbody>
</table>

We would like to introduce an index to find a part in a timetable we have to modify and to know how we should modify it in order to effectively make the timetable robust. From above discussions, it is now clear that we have to consider

- How often a delay is occurring for trains?
- How big the delays are?
- How seriously the delays are bringing knock on delays?

4.2 Introduction of a Buffer Index

A Buffer Index is an index which indicates how much delay is caused for the \( n \)-th succeeding train at the \( m \)-th station ahead for a given value and frequency of a delay of Train X at Station S. We denote this as \( B_l^{m,n} \). A Buffer Index is also defined for a particular station as \( B_l^S \) (S is the station we are interested in, such as a terminal station).

A Buffer Index is defined as follows:

\[
B_l^{m,n} = \frac{\text{Delay of the } n\text{-th train}}{\sum \text{Buffer Time}}
\]  

In (1), “Delay of the train” means the delay of the \( n \)-th succeeding train of Train X at the \( m \)-th station.
Challenge F: Even more trains even more on time

ahead. “∑Buffer Time” means the summation of all the buffer times between Train X and the n-th succeeding train between Station X and the m-th station ahead.

One of the characteristics of the Buffer Index is that we are using the delay in its definition. Generally speaking, in railways where trains are running very densely, it is almost unrealistic to dream about operating trains perfectly on time every day. We introduce the Buffer Index in order not to completely erase delays but to guess how serious a delay affects the succeeding trains. “Delay” in formula (1) might be an average or median of delays for a certain extended period of time. Thus, we can also consider frequency of occurrence of the delay in the Buffer Index.

We can regard the Buffer Index as an index which indicates a property of a delay.

In the remaining of this paper, we only discuss a Buffer Index with m=0 and n=1 (we simply denote BI₀₁ as BI). BI is an index which indicates when a train is delayed at a station, how serious knock on delays are caused to the next train at the station. BI for a train T at station S is defined as follows:

\[
BI = \frac{(Actual \ Departure \ Delay \ of \ T \ at \ S) - (Actual \ Delay \ of \ the \ Preceding \ Train)}{(Planned \ Arrival \ Time \ of \ the \ Train \ following \ T \ at \ S - Planned \ Departure \ Time \ of \ T \ at \ S) - \ minimum \ headway)}
\]

(2)

Please note that by the term “minimum headway” we mean the minimum time to be kept between the departure time of a train and the arrival time of the succeeding train.

We prepare one more definition of BI as follows:

\[
BI = \frac{(Actual \ Departure \ Delay \ of \ T \ at \ S) - (Actual \ Delay \ of \ the \ Preceding \ Train)}{(Planned \ Arrival \ Time \ of \ the \ Train \ following \ T \ at \ S - Planned \ Departure \ Time \ of \ T \ at \ S) - \ minimum \ headway)}
\]

(3)

We use BI defined by Formula (3) when we want to eliminate the influence of preceding trains. Namely, in case that we can somehow succeed in erasing delays of all the preceding trains, we should eliminate the delays imposed by the preceding trains.

In Figure 8, we show an example to calculate BI. We assume that in the planned timetable, the departure time of Train 01 at Station Y is 10:00:00, the arrival time of Train 03 at Station Y is 10:01:25 and the minimum headway between trains at Station Y is 80 seconds. We also assume that delay of Train 01 at Station Y is observed to be 30 seconds. In this case, BI based on Formula (2) is calculated as BI=30/(85-80)=6.

In other words, if we can somehow make the Buffer Index less than 1.0 for a train at a station, its delay does not propagate to the succeeding train (no knock on delay occurs). On the contrary, if a train's Buffer Index is large at a station, this means that the delay of the train at the station is seriously affecting the succeeding train. If it is impossible to make all the Buffer Indices less than 0, it is desirable to make them as small as possible.

We calculate Buffer Indices for all the trains and all the stations and we can estimate the following:

- If there is a train whose Buffer Index is by far larger than the surrounding trains at a station, it is quite likely that a delay rapidly propagates to the succeeding trains from that train at the station. In other words, we can identify a bottleneck of robustness for a timetable by focusing on values of the Buffer Index.
- In order to make a more robust timetable by adjusting departure/arrival times of trains, we should do the following for trains whose Buffer Index is comparatively large:
  - To make the numerator smaller (to reduce the delay) by making the dwell time longer, making...
running times longer and so on.
- To make the denominator larger (to increase the buffer time which becomes a buffer for delay propagation) by making the arrival of the succeeding train later and so on.

One very important thing we have to be aware of is that we must not simply put off the arrival/departure of trains. Otherwise we might not be able to set up necessary number of trains within a certain period of time (for example, in TOZAI line, it is required to set up 27 trains per hour during the morning rush hours) and as the result these trains are very congested and might cause much serious delays.

5 An algorithm to improve Robustness of a Timetable based on Monte Carlo Simulation

5.1 Basic Ideas

We have introduced an idea to improve robustness of a timetable focusing on the Buffer Indices. This algorithm covers all the process shown in Figure 1. Processes of Figure 1 are repeated many times to produce a more robust timetable.

In the Identify phase, we use the Buffer Index introduced in Section 4 (details are shown in 5.2).

In the Propose phase, we introduce an algorithm to modify dwell times of trains (described in 5.3).

In the Predict phase, we propose to use the Monte Carlo simulation method in order to estimate primary and secondary delays. We call our simulation method a fact based simulation (described in 5.4 - 6). A fact based simulation is similar to a PERT based simulation but we consider weights of arcs as probabilistic variables. We define probabilistic distribution functions for these variables from analysis of train traffic record data. We also give a probabilistic distribution for initial delays.

In the Evaluate phase, we calculate expectations of total delays for all the trains at all stations (described in 5.7).

We show the overall construction of our algorithm in Figure 9. Our algorithm works as follows:
Step 1: Conduct a Monte Carlo Simulation for a given timetable and initial delays using the fact based simulation to estimate primary and knock on delays.
Step 2: Calculate the Buffer Indices.
Step 3: Modify the timetable focusing on the Buffer Indices.

5.2 Calculation of the Buffer Index

Using the values of delays estimated by the Monte Carlo Simulation, we calculate the Buffer Indices for each train at each station. In this process, we use the Buffer Index defined by formula (2) instead of formula (3), because in the algorithm introduced in the next section, we try to fix the departure/arrival times of trains in chronological order (i.e. from morning to evening) and if an influence of earlier trains still remains, we should think the influence is inevitable and should be included in the calculation of the Buffer Indices.

Step 1 to Step 3 are repeated until a time limit is reached.

5.3 Algorithm to modify the timetable

We show an outline of our algorithm to modify the given timetable.
Step 1: Find a train and a station whose Buffer Index is greater than a certain value.
Step 2: Modify the dwell time of the train at the station found in Step 1. Its departure time is modified so that the dwell time becomes the original dwell time + buffer time /2.
Step 3: Apply PERT based simulation so that the modification in Step 2 is propagated to other trains and a consistent timetable which satisfies all the physical constraints is produced.
5.4 Fact based simulation

We have been using a PERT based simulation (sometimes called the longest path method) until now. In PERT based simulation, as shown in Figure 10, a timetable is expressed by an acyclic directed graph in which an event (a departure or an arrival of a train) is expressed by a node and the chronological dependency between two nodes is expressed by an arc. The minimum duration between two events is expressed by the weight of an arc. By calculating the longest path from the imaginary start node, we can know the earliest occurring time for an event [ABE86].

Characteristics of PERT based simulation are:
- It works very fast.
- By using an arc which indicates utilization order of facilities, we can easily express capacity constraint. Thus, we can obtain a simulation result in which all the physical constraints are satisfied.
- By introducing a pair of disjunctive arcs, we can get a simulation result in which crossover conflicts disappear or departing orders of trains are changed [TOMII95].

On the other hand, because the PERT based simulator aims at simulating train traffic assuming ideal running conditions, simulated results do not always match with actual results. Major reasons are as follows:
- Running times: actual running times of a train between stations are different from day to day according to operating conditions (delay, rain, delay of the preceding train and so on). Even for the same type of trains, actual running times may be different if the tracks at stations are different for example. In PERT based simulation, it is common to use the minimum running times which are obtained from the speed profiles of trains, but from the reasons described above, trains’ running times are not necessary the same as the minimum running times.
- Dwell times: As described earlier, if more passengers than expected get on/off a train, dwell times become longer. But PERT based simulator cannot simulate this phenomenon.
- Early departure, early arrival: In reality, trains arrive or depart earlier than planned times, although this does not happen so often. It might be critical to exactly simulate early arrival because if a train arrives earlier, trains can depart on time even when there is a lot of getting on/off. PERT based simulation usually cannot simulate early arrival and departure, because it calculates the longest path considering the planned time as well.

We propose a simulation method which we call a fact based simulation to overcome these drawbacks of the PERT based simulation.

It is now clear that we have to decide the weights of arcs of the PERT network considering the followings:
- Possibility of occurrence of primary delay
- Possibility to regain delays
- Possibility of early arrival/departure

In the fact based simulator, we treat weights of arcs as probabilistic variables and we estimate their probabilistic distributions from analysis of train traffic record data. In the next section, we describe the process to estimate the probabilistic distributions.

Fig. 10: An Example of a PERT Network (weights of arcs are not shown)
5.5 Analysis of train traffic record data

(1) Running times

We show three results of analysis about the running times of trains in Figures 11, 12 and 13. These are the scatter graphs for trains’ departure delay at a station and their arrival delay at the next station for several months.

In Figure 11, we can find a strong correlation between departure delays and arrival delays. This means that running times of these trains are the same every day. This kind of pattern was found for many stations.

In Figure 12, apparently, there are two groups. This pattern was found for trains which depart from terminal stations. We have analyzed the result in more detail and learned that trains in different groups use a different track in the terminal station. An important point is that the same running times are given in the timetable for trains of both groups. Namely, if we use the running times as the weights of arcs in the PERT network, we cannot get exact simulation results and this proves the necessity of a fact based simulation.

In Figure 13, it seems that there is no explicit relationship. This pattern was found for trains which arrive at a terminal station. But from a more detailed analysis, we found that the running times of trains in the triangle was prolonged due to capacity constraints (the track in the next station was occupied). In other words, these trains suffered from knock on delays.

From these analyses, we have reached the conclusion that we can treat running times of trains as deterministic values (not probabilistic). But it also revealed that difference of running times caused by different tracks and knock on delays have to be captured in the process of the simulation.

Fig. 11: Scatter graph for delays (strong correlation)  
Fig. 12: Scatter graph for delays (two groups)  
Fig. 13: Capacity Constraints  
Fig. 14: Dwell Time analysis
(2) Dwell times

In Figure 14, we show a scatter graph to analyze a relationship between intervals between a train and the preceding train and its dwell times. This kind of pattern can be found everywhere. Although [MEER10] reported there is a relation between a delay and a dwell time in their data, we cannot conclude that there is a correlation between the interval and the dwell time in our data. We have also analyzed relationship between delays and dwell times. But again, we could not find any explicit relationship.

So, we came to the conclusion that we treat dwell times as probabilistic variables and we define their probabilistic distributions from frequency graph of dwell times using polynomial regression.

5.6 Monte Carlo Simulation

We treat the dwell times as probabilistic variables as described in the previous section. We also treat initial delays as probabilistic variables with a distribution defined according to train traffic record data. After conducting the fact base simulation repeatedly, we calculate expectations of delays for each train.

5.7 Evaluation function of robustness

There are many discussions about robustness criteria [CAREY99]. We simply use the total sum of expectations of delays for all trains at all stations as an evaluation function of robustness. This is because the target railway line we consider is one in an urban area where trains are operated densely and passengers are expected to uniformly get on and off at every stations.

5.8 Numerical Results

We have implemented the algorithm and conducted numerical experiments using real world data. We used the timetable shown in Figure 7 as an input. Properties of the target railway line are shown in Table 2.

We included one initial delay which was known by the railway company as one of the most influential ones. We got the probability distribution of the initial delay from the train traffic record data for four months (from Sep. 2010 to Jan. 2011). We defined the running times from the same data using linear regression. We calculated the probability distribution for dwell times using the same data. Iteration of fact base simulation in Monte Carlo simulation was 2,500. We conducted modification of the timetable 25 times.

We modified trains which depart after the time of the initial delay (it occurs around 8 o’clock). The range of the simulation is from the first train to 11 o’clock, because delays usually disappear before that time.

Results are shown in Table 3.

5.9 Discussions

From Table 3, we can know that the expectations of total delays were reduced by 1,420 seconds. So, we may well conclude our algorithm succeeded to make the given time table more robust for the initial delays we assumed.

The result, however, was not so remarkable as we first expected. This is partly because the timetable was already robust enough as can be seen in Figure 7. This is also partly because we restricted the range of changes for dwell times. We were afraid that we could not set up a sufficient number of trains during rush hours (we assumed that the same number of trains as the original timetable is necessary). So, the modifications were not so drastic.

One more thing we have to mention is that our modification rule for timetable might not be acceptable
for railway companies. Currently, the timetable is made with a minimum unit of five seconds whereas in our algorithm, we assumed that the timetable can be made with a minimum unit of one second. Although there are railway companies that make their timetable with a minimum unit of one second, this might not be generally accepted and needs more discussions.

Our future work should be:
- To improve the timetable modification algorithm so that it allows other kinds of modifications such as change of pass/stop, change of departing orders of trains etc.
- To increase accuracy of the fact based simulation. In particular, estimation of dwell times needs more detailed analysis.

6 Conclusions
We have proposed an algorithm to obtain a more robust timetable by finely tuning a given timetable. We firstly proposed several methods to visualize train traffic record data such as the Chromatic diagram, the Delay diagram, 3D diagram and so on. These methods are practically used in railway companies now. We showed an example to show how we can overview the propagation of delays. We then proposed a Buffer Index, which represents a characteristic of a delay from a viewpoint of causing knock on delays. We finally introduced an algorithm to modify a given timetable to make it more robust according to the Buffer Indices. The algorithm is based on the Monte Carlo Simulation. As for the simulation algorithm, we introduced an algorithm which we call a fact based simulation in which durations between events are expressed as probabilistic variables with distributions defined from traffic record data. Preliminary results using real world timetable data show that our algorithm is quite promising.

Acknowledgements
The authors express their sincere thanks to the people of Tokyo Metro Co. Ltd. who provided the train traffic record data and valuable advices. This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (C)21510156.

References