Allocating railway platforms using a genetic algorithm

This item was submitted to Loughborough University's Institutional Repository by the/nnauthor.


Additional Information:

• This is a conference paper. The original publication is available at www.springerlink.com and the definitive version is available at http://dx.doi.org/10.1007/978-1-84882-983-1_33

Metadata Record: https://dspace.lboro.ac.uk/2134/6252

Version: Accepted for publication

Publisher: © Springer-Verlag

Please cite the published version.
This item was submitted to Loughborough’s Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/
Allocating Railway Platforms Using A Genetic Algorithm

M. Clarke, C. J. Hinde, M. S. Withall, T. W. Jackson, I. W. Phillips, S. Brown, and R. Watson

Abstract This paper describes an approach to automating railway station platform allocation. The system uses a Genetic Algorithm (GA) to find how a station’s resources should be allocated. Real data is used which needs to be transformed to be suitable for the automated system. Successful or ‘fit’ allocations provide a solution that meets the needs of the station schedule including platform re-occupation and various other constraints. The system associates the train data to derive the station requirements. The Genetic Algorithm is used to derive platform allocations. Finally, the system may be extended to take into account how further parameters that are external to the station have an effect on how an allocation should be applied. The system successfully allocates around 1000 trains to platforms in around 30 seconds requiring a genome of around 1000 genes to achieve this.

1 Introduction

A typical British railway station has several platforms each servicing rail-lines to and from the station. A station’s main purpose is to allow a train to arrive, stop and depart to a predetermined schedule, ensuring that the train is able to continue its journey on time. To effectively accommodate every train’s timetabling needs
many factors have to be considered, whilst adhering to very strict time and location constraints.

The purpose of this system is to create an application that can be used to carry out the task of platform allocation for a railway station timetable using industry standard data. A secondary purpose is to create a system that can be extended to employ extra constraints as they arise and also to integrate smoothly with a train timetable application reported elsewhere, and a resource allocation system yet to be fully researched.

The authors include professional train planners and the project’s aim is to enable them to perform train planning more efficiently and effectively. The authors admit that human skills are necessary to arrive at a satisfactory solution and the evolved solutions may be examined and edited by the human train planners.

1.1 Background

Research work has been under way for some years to develop and implement software that can provide much greater support and, gradually, should enable better schedules (both in terms of robustness and efficient use of resources) to be produced in less time. Bussieck et al (1997); Caprara et al (1997); Cordeau et al (1998); Ferreira (1997) provide useful summaries of these developments up to the late 1990s, covering timetable planning, crew and rolling stock scheduling, freight car routing, yard models, car management (all focused on a freight-dominated North American/Australian-style freight railway operation).

Work focused on generating timetables was limited until the last few years, and, as Carey (1994) highlighted (and this has not changed materially since), what there was tended to focus on single track railways (see Brannlund et al (1998); Higgins et al (1996); Mees (1995); Salim and Cai (1997), appropriate for North America and Australia, but of very limited relevance for typical European railways or complex Mass Transit networks, with short headways, trains every few minutes and diverging routes or connections to be maintained.

Of relevance is work looking to construct timetables so as to achieve an overall customer benefit, such as minimising passenger waiting time (see Daduna and Voss (1995)), or a combination of this and operating cost (see Chang et al (2000); Nachtegaal and Vogt (1997)). Whilst this is focused on the passenger, it does not fit very well with the developing European railway industry structure, where railway infrastructure providers need to focus on the requirements of their customers, the train operators, more than the ultimate customer, the passenger or the freight shipper.

Carey worked for some years on the generation of timetables for complex European railway networks. In Carey (1994); Carey and Lockwood (1995); Carey and Carville (2000, 2003); Carey and Crawford (2007), he describes and extends the discussion on algorithms to generate timetables, highlighting along the way the particular problem of station infrastructure complexity and considering whether this should be treated as a separate computing task. Comparable work has also been undertaken
in the Netherlands. Kroon et al (1997); Odijk (1996) providing early papers setting out work to develop algorithms for generating railway timetables; this has culminated in the development and implementation of the 'DONS' software package for Railned, the Netherlands state-owned railway infrastructure provider Hooghiemstra et al (1999).

More recently work on timetable generation has continued to emerge from Dutch universities, but focused typically on generating a 'standard hour' timetable Peeters and Kroon (2000); Liebchen (2003)) rather than the less regular type of timetable often found in the UK. The number of papers, and the complexity of the timetabling problem being investigated, have increased in the last few years, with Liebchen (2007); Ingolotti et al (2006); Rodriguez (2007); Tormos et al (2007)) describing research underway seeking to generate feasible timetables for complex European railways. The European Commission now provides a web site for researchers to share information on research under way in this area.

Also of interest is the approach adopted by London Underground through until 2008, which was the subject of a paper presented by Wallace (1995), although ‘metro’ operations have rather simpler timetabling challenges than ‘main line’ railways.

Platform allocation is currently done manually in the UK. For many locations train planners use spreadsheets or even graph paper and then transfer the solution into the train planning systems. Some software (e.g. RailSys) has the functionality to ‘drag and drop’ trains between platforms and highlights conflicts. In all cases the train planner in not given any support in terms of which platforms to use for particular trains to get an acceptable solution.

2 Problem Analysis

Allocation of station platform time to any participating train requires all the information related to the station to be known. For a traditional ‘stopping’ train information concerning its anticipated arrival time at the station, how long it requires platform time for and when it departs must be established before the allocation of a platform can take place. This must be obtained from the available industry data that are used to record and distribute train route information to all those parties that need it.

A single train route details the stations that a train will either stop at, or pass through but the data do not provide the inverse relation, to answer ‘what trains are stopping at, or passing through, station X?’ To obtain these data preprocessing must be performed providing a relation between stations and trains.

Pre-processing of the data is necessary before the allocation of platforms can take place. Analysing train data may be divided into two parts, with a further third part necessary to do the actually platform allocation. Each part acts as a pre-requisite to the next, leading to a final integrated system.
2.1 Part one: Linking trains to stations

A Common Interface Format file (CIF) Kitchin (2005), holds route data from the Network Rail Train Services Database (TSDB) in an electronic format. The data only provide details of a train’s route data, and do not directly provide anticipated station use. This information must be deduced from the data supplied.

RWA have to resolve which trains are scheduled to arrive at which stations from the CIF files. From using arrival and departure information it is possible to deduce all the trains that will stop at a specific station. Until all the anticipated trains that use a station are known, it is difficult to accurately allocate suitable platform time to all those on the timetable without any clashes.

To provide a solution to this the system must be able to process CIF data and associate every route with the stations that it is anticipated to stop at, or pass through. Once associated it is be necessary to convey this information to the user, giving them access to a station’s basic timetable so that it may be viewed before any platform allocation is performed.

2.2 Part two: Platform Allocation (Internal)

Once all trains have been linked to their stations, they need to be given an appropriate platform allocation. A station is likely to have many platforms, varying from station to station, and has to accommodate trains that arrive or depart within minutes of one another.

By using the earlier calculations the system has the timings for every train at a station, what time a train arrives, how long it is stationary for, and when it is scheduled to depart. From this calculated information it is possible to see where train routes interfere or ‘clash’ with one another.

The suitability of an allocation can be judged upon the number of these clashes, where a timetable allocation with fewest clashes is best.

There are potentially further considerations than the timing constraints of a route, with the possibility of commercial and service line requirements that need to be adhered to. A train is often able to arrive only via a certain set of lines, dictated by where it has come from, and may only leave on a limited set thus restricting the platforms it may choose Carey and Crawford (2007).

A particular rail company may want all their trains to use the same platform on a particular route, due to commercial considerations. The system needs to be able to accommodate such requirements where possible, whilst ensuring minimum disruption of train routes.
Platform Allocation

2.3 Part three: Extended Platform Allocation (External)

Further constraints on a railway station and its environment can directly or indirectly affect a train’s route to and from a station.

This increases the complexity of the system and the processes that manage the data associated with performing platform allocation. There are other problems that need to be considered, such as two trains scheduled to depart on different lines but where the lines cross at a nearby junction.

Train characteristics can also dictate where a train should be given platform time, as different trains could potentially have differing travel speeds that again could potentially lead to trains ‘catching’ each other up and cause disruption between trains external to the station. Potential clashes need to be resolved with minimal interference to either train, this is known as ‘headway’.

Introducing tighter constraints on how a train is to be allocated platform time allows the system to be able to accurately perform the allocation for the station’s timetable using complete data. Many new constraints may arise as train policy changes and the system must be capable of dealing with these.

3 Documentation analysis

There are several pieces of documentation that are currently used for the platform allocation.

3.1 Common Interface File (CIF) Document

The CIF file holds the data on the schedule for a train and all the necessary route information in a standardised format (Figure 1). The CIF file is a sequential text file consisting of fixed length records of 80 characters.

It contains various record types that are denoted by a ‘record identity’, the first two characters of the record. The record type determines the structure of the record.

The CIF is the main source of train information that will be required when carrying out platform allocation. Associating all the train data requires the processing of the majority of the data held in this file. The development will be primarily concerned with train data for the Glasgow area. In particular the main station that will be the focus of development and testing is Glasgow Central and the results presented in this paper use these data.

All necessary train information must be extracted from the CIF, and to assist in this process is the aid of a CIF End User Specification documentKitchin (2005) that details all the possible CIF data.
3.2 End User Specification Document

A typical train route will be structured with 5 forms of record type.

**BS** Basic Schedule Record
- Details the main characteristics of the train route
- Identifies a unique Train UID code to determine the route
- Information on header ID (the train header unit)
- Details on operational days, train speed and further characteristics

**BX** Basic Schedule Extra Record
- Is an extra record for the BS with a few fields to detail Continental Europe trains

The next three records of LO, LI and LT are all concerned with the stations that the train will interact with.

**LO** Origin Location Record
- Always used to denote the initiation of a train route, the station at which the route commences.

**LI** Intermediate Location Record
- Details on whether the train is stopping or passing and its requirements at a station during the route.

**LT** Terminating Location Record
- Concerned with the final station that the route will terminate at.

LO, LI and LT records hold the majority of information on train arrival and departure times. As such they are essential to performing the initial train association.

Focusing in on a single LO record (Figure 1), the majority of information regarding the departing train can be derived based upon the record descriptions given the End User Specification document (Table 1).

This extract is taken from a CIF file detailing the train routes for the Glasgow area on a single day (Wednesday 11th October 2006). This file has nearly 39000 records, illustrating the quantity of data that currently has to be processed just to satisfy a single day’s worth of timetabling for a city area.
Platform Allocation

Table 1 Breakdown description of a single LO CIF record fields

<table>
<thead>
<tr>
<th>Value</th>
<th>Field Name</th>
<th>Size</th>
<th>Format</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>LO</td>
<td>Record Identity</td>
<td>2</td>
<td>A</td>
<td>origin Location Record</td>
</tr>
<tr>
<td>AYRR</td>
<td>Location</td>
<td>8</td>
<td>A</td>
<td>TIPLOC- unique station code</td>
</tr>
<tr>
<td>1810</td>
<td>Scheduled Arrival</td>
<td>5</td>
<td>A</td>
<td>24-hour format</td>
</tr>
<tr>
<td>1810</td>
<td>Public Departure</td>
<td>4</td>
<td>N</td>
<td>24-hour format(unpopulated)</td>
</tr>
<tr>
<td>4</td>
<td>Platform</td>
<td>3</td>
<td>A</td>
<td>Platform intended to be used</td>
</tr>
<tr>
<td>TB</td>
<td>Activity</td>
<td>12</td>
<td>A</td>
<td>‘TB’ (Train Begins- mandatory value)</td>
</tr>
</tbody>
</table>

3.3 Docking Movement File

A final document is a spreadsheet detailing the train movements for a day at a single station. Movements of trains, or header movement, occur typically at a station where a terminating train is then used to form the train for a later originating route. The file is a listing of all the movements at a station whilst detailing the expected station use of any arriving train such as what line and platform the train should be using.

Also data on the process of station switching is detailed where a train may need to move from one platform to another in order to continue on its route, as it may now need to switch lines in order to reach its next destination.

Information regarding this needs to be used to handle constraints concerning external requirements for a station’s platform allocation. For it to be useful the information concerning trains needs to be linked in some form to the already existing data extracted from the CIF file.

4 System Specification and Design

Section 2 discussed the need to associate the train data and establish the quantity and type of data that needs to be included, thus allowing platform time to be distributed more precisely.

The high-level design illustrates the need for a separate association module that allows all the necessary pre-processing to be performed before a Genetic Algorithm Holland (1975) is implemented. Two input files will be required to implement a train association function, the CIF file containing the train data and a configuration file.
4.1 Specification

4.1.1 Linking trains to stations

This creates a data structure of station timetables all correctly associated with their entire train schedule for the day. Records are read in from the CIF sequentially, establishing the train information from the BS record that precedes the stop information of a train route. Each stop record is then processed establishing the station, stop timings and any further details. These are then added to the relevant station schedule, its insertion position determined by its time value. Having the final station schedule in an already sorted order allows easier execution of the Genetic Algorithm and fitness calculation for platform allocations. Once all CIF records have been processed every stop record should be associated to a station schedule, giving all station timetables that can now be in the position to perform platform allocation.

4.1.2 Genetic Algorithm

The appropriate station timetable is selected for allocation.

Specifying the Genetic Algorithm parameters state the structure of execution for the algorithm, whilst the input of a Docking Movement File, Section 3.3, will give details of any external considerations that the allocation module will have to consider.

An initial population of candidates is generated at random adhering to the chromosome structure that is to be operated upon. Each initial candidate in the population has a fitness value assigned to determine its suitability as a plausible solution Floreano and Mattiussi (2008). Breeding phases are then executed round by round, generating new candidates through the use of pre-determined operators. At each round the population fitness for every member is recalculated and the process of selection is started again for the next generation. This continues until either a fit solution is achieved or the round limit has been reached.

As with associated train data, a single station timetable can be stored for intended future use. Past single station timetables can also be restored when wanting to just perform an allocation on a known station. Once achieved with the use of the Genetic Algorithm the solution can be stored permanently on the file system. Likewise previously generated allocations can be restored through the system.

There is the ability to amend already processed schedules by allocation that allows a single train record platform allocation to be altered individually. The fitness can then be recalculated to evaluate the suitability of this amended candidate.
4.2 Store and Loading Data File Design

In addition to being able to perform the necessary processing of train data and achieving platform allocation solutions, there will be the need to have functionality that will allow the permanent storage and retrieval of outputted data.

Associated data, a single station timetable schedule and single allocation candidates have all been identified as those data structures that a user of the system may wish to store digitally. Likewise a facility then must be in place to make use of previously saved data that may need further processing by a user.

These files that are formed from storing the data must be designed to a particular format so that the system is able to distinguish them apart from one another when coming to restoring them back to the system.

5 Genetic Algorithm Design

There are to be several determining factors that have to be processed by the allocation algorithm when implementing the Genetic Algorithm.

5.1 Timings

Arrival and depart times at the station is the primary factor to distributing a station platform to a train record.

These times are used to ensure that when trains arrive, are stationary and depart a station at their scheduled platform, no other trains will clash with them.

A clash can occur between two trains A and B when

- Train A does not depart from a station platform before a later train B arrives at the same platform.
- Train A does not depart from a station platform before a later train B requires the same platform to depart from (when train B is the commencement of a train route).

Using this as the main factor for deriving the suitability of a candidate for platform allocation, an initial Genetic Algorithm structure can be devised along with the required chromosome composition.
5.2 Genetic Algorithm Structure

To form the basis of the Genetic Algorithm, the process to schedule platforms for trains that are to arrive at a single station using a simplified platform structure will follow:

- Each candidate solution comprises of a timetable schedule established by associating all train data.
- The population is formed from many of these timetables with their allocations.
- Each timetable lists the unique trainUID (acting as the identifier for that train route) and the platform that it has been allocated. (Table 2)
- A train record holds the necessary information on the train, and its parameters used to decide on the fitness of the proposed allocation. (Table 2)
- The fitness function is applied to each genome initially to test for the number of clashes that occur across the platforms allocations applied.
- The higher the number of clashes the more unfit the solution and thus will be less likely to be included in the next generation of candidates.
- A candidate that has a smaller number of clashes will have a stronger fitness and be more likely to be selected for the next population for breeding.
- The size of the chromosome is determined by the number of train routes that are to be scheduled at the station in question.
- If there are 20 trains to be allocated then each chromosome will comprise of separate 20 genes (train records).

5.3 Chromosome Structure

<table>
<thead>
<tr>
<th>Candidate A</th>
<th>Arrival Time</th>
<th>Depart Time</th>
<th>Platform</th>
<th>Clash?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train A</td>
<td>1102</td>
<td>1105</td>
<td>1</td>
<td>Y</td>
</tr>
<tr>
<td>Train B</td>
<td>1104</td>
<td>1107</td>
<td>1</td>
<td>Y</td>
</tr>
<tr>
<td>Train C</td>
<td>1108</td>
<td>1115</td>
<td>3</td>
<td>Y</td>
</tr>
<tr>
<td>Train D</td>
<td>1109</td>
<td>1110</td>
<td>2</td>
<td>N</td>
</tr>
<tr>
<td>Train E</td>
<td>1110</td>
<td>1112</td>
<td>3</td>
<td>Y</td>
</tr>
<tr>
<td>Train F</td>
<td>1115</td>
<td>1116</td>
<td>2</td>
<td>N</td>
</tr>
<tr>
<td>Train G</td>
<td>1115</td>
<td>1117</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>Train H</td>
<td>1119</td>
<td>1121</td>
<td>4</td>
<td>Y</td>
</tr>
<tr>
<td>Train I</td>
<td>1120</td>
<td>1120</td>
<td>4</td>
<td>Y</td>
</tr>
<tr>
<td>Train J</td>
<td>1126</td>
<td>1128</td>
<td>4</td>
<td>N</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Platforms</th>
<th>Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2,3,4</td>
<td>3</td>
</tr>
</tbody>
</table>

The chromosome used is the Platform number.
5.4 Genetic Algorithm Parameters

Initiation of the algorithm requires several predefined parameters that distinguish the behaviour of the search.

- **Population Size**: the number of potential candidates that must exist before any genetic operations can be applied.
- **Tournaments**: the number of breeding phases that the algorithm is limited to perform until a fit solution is achieved.
- **Fitness Target**: the target for the algorithm, should a candidate achieve this value the current round should finish execution and the algorithm should stop. A zero (0) target is generally expected but others may be defined.
- **Reproduction Rate**: the proportion of candidates generated through reproduction in the breeding phase as a percentage of the total population.
- **Crossover Rate**: similar to the Reproduction Rate, but denotes the proportion of candidates generated for the population via the Crossover operator.\(^1\)
- **Mutation Rate**: A percentage value of the possibility of mutation occurring, usually small.

\(^1\) Reproduction Rate and Crossover Rate are required to sum to 100% so to generate the necessary population value.
6 System Output

The system has a GUI that enables the user to view the results, for example Train to Station linking, GA Execution Output, Platform Allocation Results (Figure 2).

The system interface is very flexible. For example, if there is more than one candidate solution then the user can step through them, which allows the train planners to explore possibilities.

6.1 Fitness Function

The fitness is simply the number of clashes in the schedule. However, some stations have hundreds of trains on their schedule for a single day, calculating a fitness of a candidate needs to be efficient.

For the originally designed fitness function design, if a station had 1000 trains to allocate, then an algorithm comparing station and timing constraints for every train with every other train a population of 100 would require between 50 million and a billion comparisons to establish the fitness of the entire population.

To reduce the execution time and complexity of the algorithm, the fitness function was redesigned. Instead of forcing the algorithm to compare the list of all the remaining stops that occur for that day, it is more efficient to only compare those trains that the function knows will be at the station in the same time window.

If train A is to depart at time DA, then the trains that are included in the fitness function for comparison are those that will be at the station before train A departs at time DA. This gave a significant reduction in the complexity of the algorithm.

7 Conclusions

The project has succeeded in developing an application that can aid train-station platform allocation. The system is able to process the complex array of data that traditional manual methods have to handle, whilst delivering a workable method for distributing platform time to trains.

The introduction of further working files, CIF Configuration and Parameter File, to aid the task of linking the train data in the CIF (Common Interface Format) led to careful design in those file structures and how they were to be used in the system. They give more control to the user and if the structure of the CIF changes then these files can be modified to allow the system to process later versions.

The association of data puts the timetable allocator in a position of having all the station timetables at their disposal with accurate information on those schedules.

The Genetic Algorithm to allocate platform time at a station can quickly achieve an allocation with minimal clashing between scheduled trains. The system gives
Platform Allocation

feedback to the user so if it appears that an allocation will not be achieved, the parameters of the algorithm can be easily changed.

The system allows the manual alteration of allocations if a maximally fit solution is unavailable. A user can accurately specify a platform allocation that the system has failed to find. Providing the ability to automatically and manually specify platform allocation is more likely to lead a fit allocation being achieved.

The Docking Movement File has not been considered at this stage as most stations do not require this, so this system has addressed a large part of the problem.

This project met and addressed the requirements of the first two problems outlined in the analysis. It has successfully used a Genetic Algorithm to perform platform allocation for a station in a generalised manner. The genome used is very large compared to other problems tackled using a genetic algorithms, however the large search space is populated very heavily with possible solutions. One requirement of train planners is for a variety of solutions to choose from. This system delivers that.

References