Multi-train simulation: verification and accuracy

William A M Barter
Comreco Rail Ltd., York, England

Abstract

Many simulation packages model the movement of individual trains, to establish the speed profile, or support timetable planning. Some extend the potential of simulation to model multi-train operations by adding data on the network infrastructure.

Extra functionality is achieved by increasing the range of input factors. A single train simulator will match train mass with tractive effort to calculate train acceleration. Multi-train simulation needs detail of signal section lengths, the track layout and locations of critical track circuits. Validity is improved by further additional input factors, such as gradients, train resistance, transmission efficiency, equipment and driver reaction times, and specific signalling rules.

However, comprehensiveness brings difficulties. A package capable of detailed simulation will be complex and data-hungry. Operators must be trained - their knowledge must be used regularly or lost. Model-building will be expensive, and elapsed time possibly prohibitive in practice. Input errors may obscure genuine effects. Preparing and using the model also needs knowledge of traction, signalling and interlocking, and understanding of timetabling. The skills will rarely be found in one individual, but require a pool of users, increasing the ongoing workload needed to support the activity.

Moreover, each data item brings with it some uncertainty. Where new input opportunities extend into data areas that are poorly researched or intangible, values may be estimated or guessed, and added uncertainty mask the added validity. This is particularly true where human behaviour is involved.

Network simulators are ideal for specific and well-defined tasks, but for “high level” applications detailed input is immaterial. Comreco Rail Ltd. has risen to the challenge of offering a simulator that minimises the workload input, avoids
reliance on specialists, and offers results within the timescales of management decisions, by adding simulation functionality to the \textit{TrainPlan} timetable management system.

1 \textbf{Introduction}

Simulation - a representation of a system, capable of accepting data on the major factors affecting its behaviour, and able to predict future or unprecedented states of the system, by means of defined relationships between the various input factors – is now a widely accepted tool for planning railway infrastructure, traction and train services.

Such a simulation model will be characterised by an initial “setting-up” time that is high compared with the time then required to define a scenario, run the model and obtain results. The user actions should be small compared with the volume of calculation – this in practice implies that the simulation will be computer-based. This contrasts with manual systems, such as traditional timetable graphing, which need minimal setting up but incur very high “operating” times.

2 \textbf{Simulation and railways}

Many software packages are available which will calculate station-to-station running times for specimen or typical trips. Such a package is termed a “Single-train Simulator”, or “Train Performance Calculator” (TPC). Essentially, the package will use the mass of a train and the tractive effort of its power unit to calculate its acceleration, on the basis of Newton’s laws of motion. Typically, tractive effort falls as speed rises, so that the acceleration calculated on starting from rest is applied for only a short time or distance before a new value for tractive effort is selected from the input data, and the calculation repeated. Further recalculation is then made at frequent intervals, and the acceleration then continues until either the train or route speed limit is reached, or until a requirement to brake is perceived.

Refinements of this fundamental operation may take into account such factors as:

- gradients;
- train brake force rather than a global rate;
- train rolling and aerodynamic resistance, possibly modified locally for curves and tunnels;
- adhesion limitations on acceleration;
- limits on rate of change of acceleration (Jerk Rate);
- changes to payload at each station.

Simulation at this level more or less accurately models the Speed : Distance behaviour of individual trains. No account is taken, however, of interactions
between trains. A multi-train simulator requires further input on the signalling and interlocking that in reality will separate trains on route and at junctions.

The simplest use of a multi-train simulator is to calculate line headways, that is, the minimum interval between trains that is permitted by the signalling. The headway reflects a complex interaction between the lengths of signal sections, station locations, train braking and acceleration, gradients, line speed and local speed restrictions.

Standard methodology for Headway assessment involves trials with multiple trains at varying frequencies, to spot the lowest interval that still avoids headway infringements, judged by whatever criterion is appropriate to the railway in question. For this analysis, a package based on a Train Performance Calculator is essential, so as to capture the precise Speed : Distance behaviour of trains, but the package must also be able to accept sufficient signalling detail as to relate this to signal locations.

Together with running times identified in single-train simulation, output on the headway from this simple multi-train modelling provides the data from which a timetable can be compiled. The number of trains that can be integrated into a conflict-free timetable, taking account of the traffic mix, junction conflicts and terminal working, is in fact quite a fair definition of the capacity of a network. Preparation of a timetable, however, says nothing about the reliability with which the planned service will operate.

A “correct” timetable will be “conflict-free”, implying that, if all trains start on time and adhere to the planned running times, there will be no signal checks at junctions or on route. In reality, the risk is that a late-running train will present itself at a junction or at the entrance to a “corridor” simultaneously with a punctual train. If the otherwise punctual train is delayed, as might be the case under some prioritisation strategies, the system now contains two late trains, each of which can then delay others at another junction. Alternatively, the late train might be further detained - either way, the total delay increases.

Clearly, the risk of escalating delay increases as the number of paths used rises compared with the number of paths available. Any experienced operator or train planner will have a feeling for this robustness. But from timetable compilation alone, no quantified conclusions can be drawn as to the robustness of a timetable.

It is here that, with the multi-train capability, simulation comes into its own. Given detail of the track layout and interlocking, the crucial step is to enter a probability distribution for delays to trains at the start of their simulated journeys. The simulator samples the distribution whenever a trip is initiated, so that a proportion of trains starts late. This technique is known as Monte Carlo simulation.

During the simulation run, the action of the signalling interlocking will prevent conflicting trains from occupying a junction simultaneously, leading to additional delay to one train or the other. The difference in additional delay between scenarios indicates the benefit or otherwise of the differences in their components.
As an example, in a simulation experiment [1] based on the three alternative layouts for a small station, the following figures were obtained when the same timetable, conflict-free for all the layout options, was modelled:

- Layout 1 – 0.56 minutes additional delay per train;
- Layout 2 – 0.75 minutes additional delay per train;
- Layout 3 – 0.52 minutes additional delay per train.

From this it can be concluded that layout 3 is preferable, and that if any extra capital cost is involved, it may be justified by the saving in delay. The actual expenditure that can be justified by avoiding delay will depend upon the revenue earned by the trains in question.

3 The costs of simulation

So simulation is an effective tool for calculating running times, assessing line headways, and looking at the detailed behaviour of networks. At this point, we might take stock and consider the implications for any organisation wanting to exploit simulation.

A package capable of detailed multi-train simulation will be complex and data-hungry. Operators must be trained, and their knowledge must be used regularly or lost. Preparing and using the model also needs knowledge of traction, signalling and interlocking, and understanding of timetabling. The skills will rarely be found in one individual, but require a pool of users, increasing the ongoing workload needed to support the activity.

Purchase of a simulation package probably involves a licence price that is significant for typical railway administrations, especially those in the public sector. Moreover, building detailed models is expensive, and the required elapsed time may be out of line with the timescales of management and business decisions.

The workload of traditional simulation is exacerbated by the fact that although a particular study may use only a fraction of the functionality of the package, much of the underlying database has to be populated if the package is to operate at all. Ironically, as use of a simulation moves into the higher-level network functions, the most basic data starts to become of reduced significance. If the main impact on the results arises from interactions between trains, with trains possibly waiting minutes for a path across a junction, then effects arising from the detail of the traction data, probably measured in seconds, are lost in the overall limits of error. Time may be saved by using estimated or generic data for less important input factors, but if this is acceptable would it not be better to avoid the need for such input data completely?

Therefore, a railway wishing to exploit simulation is faced either with a significant investment of capital and labour, or with the ongoing expense of using consultants. And what does it get for its money?
4 Are simulation outputs useful?

The output figures for additional delay quoted above are small, and the differences between them even smaller. Are they demonstrating real effects and valid differences? To put this in context, remember that Railtrack’s average delay per train movement (from the 1999 Network Management Statement) is 79 seconds. Clearly the small figures above can represent a significant part of the underlying network punctuality.

But one very clear risk with a complex and detailed model is that of input errors. Having argued that small figures may nevertheless be significant, perhaps I can put this in perspective against the results of the simulation:

- Layout 1 – 0.56 minutes additional delay per train;
- Layout 2 – 0.75 minutes additional delay per train;
- Layout 3 – 0.52 minutes additional delay per train
- Layout 2 with input error – 1.84 minutes additional delay per train.

The error in question concerns release of the overlap of a platform starting signal after arrival of a terminating train – one that requires a fair knowledge both of signalling practice and of the specific simulation package to spot and correct. Were the error repeated in all scenarios, and the results then taken at face value, it might easily be concluded that the differences were not significant.

All these deficiencies are compounded as the geographical area to be modelled expands. The impact of optimistic output may compound. The more data there is, the greater is the risk of errors first being made, and then going unnoticed, especially when the work is being performed to a demanding budget and timescale.

Anecdotally, I suggest that most simulation results are optimistic. Even for relatively straightforward issues such as headway assessment, it appears that the real operators are reluctant to place too much reliance on our results. The theoretical headway with 4-aspect signalling for 125 mph is around 92 seconds [2], and simulation of parts of the UK East Coast Main Line have shown that the actual headway is very close to this ideal. But the minimum headway on which the timetable is based is 3 minutes – nearly twice the theoretical figure. Clearly those who plan and run the railway are not content to base their operation on simulation output without very considerable contingency factors. The margin added to ensure operability is equal to the whole base figure derived from simulation.

There are reasons for this, of which caution need not be the most significant. Obviously, some allowance for reliability is needed, to ensure that a trivial delay, or slight variation in traction performance or driver behaviour does not escalate into significant delay to following trains. One factor proving to be of significance especially with solid-state interlockings, especially those remote from the control centre, is equipment response times deriving from speed of data transmission.

5 Improving simulation validity

The obvious route to improving validity is to add further input factors, such as data on transmission and braking efficiency, equipment and driver reaction
times, and specific signalling rules. However, each data item brings with it some uncertainty – the more obscure the data, the greater will be the uncertainty relative to the incremental validity. And a data entry opportunity is useless without data to be entered - where new input opportunities extend into areas that are poorly researched or intangible, values may have to be estimated or guessed, which is quite pointless.

Increasingly we find that the gaps in our models are in areas in which human behaviour is a factor, perhaps train braking or station dwell times. In such areas we are not talking about discrete values, but rather a range of possible values occurring with varying frequencies. Stochastic Monte Carlo techniques are fully appropriate in such cases, but to support such an analysis a large amount of data on the factors in question must be gathered. I question first whether the validity that would be added to the simulation outweighs the uncertainty inherent in this information, and second whether the workload involved is commercially realistic, unless the issue is to be the subject of the study.

Then, do our outputs actually answer the right questions? This is more an issue for study design than system functionality, but is it not tempting to offer our clients what our systems can do, rather than try to establish what the client wants? Sometimes this is just a matter of selecting appropriate outputs for reporting from the range offered. Numbers of restrictive aspects and average delay per train are both measures of system reliability, but it is useless to offer the former to an investment analyst and the latter to a signal engineer. The UK has extremely strict penalty regimes in place between the participants in the industry, but Railtrack’s performance is measured in minutes delay whilst train operators are judged on the percentage of trains on time and within specified delay criteria. It has been found that action that improves one measure may lead to deterioration in the other, so the outputs reported must match those relevant to the particular client. Sometimes the problem is in specifying the question that the simulation is to answer, but more often the figures that the simulator produces simply do not form an input to the next stage of the business decision-making process.

Moreover, because of the missing factors outlined earlier, simulation output should probably not be treated as correct in an absolute sense, but used simply to compare different scenarios. When looking at a change to an existing railway, present operations can be modelled as a benchmark, but at the expense of adding to the workload and delaying the results the client wants. Benchmarking is not possible at all for a new system. It may be argued that any simulation is better than none for a new system, as some uncertainty is removed, but the more likely impact is to give an unjustified credibility to a false, and probably optimistic, result.

Over the last decade or so we have seen TPC-driven network simulators extended so as to try to cover the whole range of applications outlined in this paper. I suspect we have gone as far, and probably further, down this road than is sensible.
6  So where next?

I believe that the future will drive commercial suppliers of simulations, and their customers, in three different directions. First, the most tangible uses of traditional simulators - train performance calculation and headway analysis will be maintained. The challenge to the supplier and user is to reduce the data input workload by “specifying down” the models precisely to their intended purpose. There is scope for additional functionality to improve validity – examples are equipment reaction times, more sophisticated representation of train braking as drivers really do it.

Comprehensive network simulations will continue to be of value for investigating specific issues within limited areas. Other valid applications of such detailed simulations include investigation of the impact of technical failures, and calculation of demands on the power supply system, including the impact of signals checks, which can increase energy consumption significantly compared with free-flow. In the former case detail of the equipment is essential so as to represent the relevant failure, and in the latter the detail of the traction is vital.

Secondly, the alternative approach to comprehensive simulation is to rely on train planning to analyse capacity. A network simulation requires a timetable as an input, and I have often observed that much of the benefit of the simulation in terms of suggested courses of action and practicable conclusions has resulted from the task of preparing that input rather than from what the simulator does with it. In projects major and minor, the relatively simple task of preparing a timetable, either in its own right or as an input to a simulation, is overlooked and underrated.

Finally, for high-level tasks such as analysing reliability of extensive networks, perhaps mimicking performance regimes to allocate penalties, I believe the right approach lies in increasing the intelligence of the timetable planning system.

7  An intelligent timetable compiler

At VST Comreco Rail Ltd., we therefore set ourselves of devising a network simulator that would

- require the minimum data consistent with validity;
- use only readily available data;
- operate in a standard train-planning platform, so as to be available to the railways regular staff rather than depending upon a specialist input;
- permit analysis of perturbed operations, taking account not only of interactions between trains at points of conflict, but also of associations between trains arising from stock and crew diagrams leading to propagation of delays;
- offer quantified output that relates to business and management decisions.
Our TrainPlan timetable compilation system is in widespread use amongst operators in the UK and abroad, and was identified as the basis for such a set of simulation functionality.

TrainPlan is based on a database of Sectional Running Times (SRT’s), catering for each relevant combination of Train type (known in UK terminology as a Timing Load), and Station to Station pair (known as a Network Link). Different SRT’s will apply, depending upon whether the train in question is to stop at both stations of the Network Link, or one, or neither.

The SRT’s are user-defined, but might be derived from current knowledge or running trials, or from a Traction Performance simulation. The fact that the user is required to authorise each SRT rather than simply drawing it from another system allows an element to be added to represent operating realities.

TrainPlan also holds data on the infrastructure, essentially relationships between Network Links, including information on points of conflict within the track layout which imply that use of one network link prevents simultaneous use of another, and the headway for each Network Link, that is, the minimum interval between trains on entering the link or on leaving it. The Entry and Exit headways will be specific to different types of trains, and may differ where trains run through the Network Link at different speeds. Where Network Links conflict, the interval between successive occupations is stored.

The basic functionality of TrainPlan allows the user to compile a trip schedule by entering one key time (normally the required starting time for the trip, but quite possibly the required arrival time, or time at an intermediate station), and then defining whether the trip is to stop at or pass each station on the route. By drawing on the database of SRT’s, TrainPlan will then derive actual arrival and departure or passing times for all other stations.

Having compiled a number of trips in this way, the user is then able to initiate a check for conflicts between them, both on route and at junctions, by comparing actual timings of trains and checking the intervals between trains at key points against the permissible margins stored in the geography database.

A timetable thus prepared in TrainPlan provides a basis for assessment of vehicle and crew requirements and compilation of the vehicle and crew “diagrams”. This function also requires storage of data on factors such as minimum turnaround times between trips worked by the same set of vehicles. Other associations between trains, such as connections, are also stored.

The basic functionality outlined above fails to allow the forecast timings of one train to modify those of others, as the system simply identifies where planning margins are infringed by the planned timings. This in reality or in a true simulation would lead to an interaction between trains. It is true to say, however, that a timetable plan for which no infringement warnings are generated is as valid a demonstration of a conflict-free plan as a traditional simulation that runs without generating delays to trains.

It is a logically a very short step from identifying that two trains are planned to be at the same location within a given time to altering the timing of one (normally the later of the two) to be equal to the time of the first plus the
minimum time interval for the train types and the location. This short step, however, is the critical difference between a single-train and a multi-train simulator.

Current functionality developments in TrainPlan were designed to automate timetable compilation and modification, for instance, by identifying vacant paths. In this function, a vacant path would be identified by:

- Initiation of a trip as early as possible after a user-defined time;
- Progressively timing that trip in accordance with the SRT’s appropriate to its type, route and stopping pattern;
- Identifying any infringement of planning margins at each location;
- Adjusting the planned timing for the new trip on the basis of the minimum acceptable margin between conflicting trains.

On completion of a trip schedule, the equivalent process is operated in reverse, by taking the arrival time achieved by “forward timing” and working back towards the origin (“back-timing”). This serves as a check as to whether that arrival time could have been achieved by a later departure. If so, the path associated with the later departure is accepted.

One further facility in TrainPlan is an optimiser for single lines, working within user-defined constraints to path trains between crossing points with minimum loss of capacity.

8 From timetable to simulation

The first three of our criteria – requiring minimum data input, using readily available data, within a standard platform – were therefore met by production versions of TrainPlan. In producing a trip schedule from a minimum of input data according to pre-defined rules, TrainPlan just falls within the definition of a simulator of a single train nature, the difference being that running times between stations are pre-defined, rather than being calculated as part of the operation. The system is able to modify a planned timing on the basis of an identified conflict, providing a basis for modelling interactions between trains.

The key element that turns a timetable compilation system into a true simulator is Monte Carlo simulation, that is, randomised representation of variable events. In the case of rail networks, the normal use of randomisation is to add delays to trains at their entrance to the simulation area. These “Start Delays” are chosen from a user-defined probability distribution. The missing functionality relates to ability to model perturbations – essentially a matter of defining appropriate input opportunities – and to the system outputs.

With the facilities already available within TrainPlan, this simple addition will:

- cause a certain proportion of trains to start late;
- for each delayed train, adjust the forecast arrival at the next location to reflect the Start Delay;
- identify conflicts with other trains encountered whilst running to that location;
• adjust the timings of either of the conflicting trains to resolve the conflict;
• repeat the process for each further step in the journey of each train, including any trains now subject to delay as a result of conflict resolution.

The impact of conflicts created and resolved in this way will be felt in the ultimate arrival time of trains at their destinations or at the exit from the simulation area, termed the End Delay. Comparison of End Delays with the Start Delays indicates the extent of conflicts within the simulation area.

Refinement of this process may take account of factors such as train prioritisation strategies, and extension of running times where trains are stopped to avoid conflicts.

As well as randomised delays, it is also appropriate to allow the user to define specific delays, e.g. a given delay to a chosen train, or to all trains from a particular origin.

By decoupling the model from traction performance calculations, and by defining the impacts of the signalling directly rather than deriving them from detailed source data, the input workload for a model at this level is much reduced compared with a traditional network simulator. Moreover, values used for running times and planning margins can be selected to reflect the realities of actual operations rather than perpetuating ideal circumstances in the network model.

In this way the timetable planner has available within the normal planning tool the facilities to confirm how a planned timetable responds to delays, and, by controlling the delays imposed, compare one timetable with another for potential reliability, examining, for instance, the impact of utilising spare paths, or of providing additional infrastructure to reduce potential for conflicts.

9 Further development

The simulation facilities outlined in this paper will be developed to reflect time penalties when trains are stopped or checked to await a path, and to incorporate more sophisticated train prioritisation strategies.

Utilisation of a train planning platform for such a simulation presents further opportunities which may not be found within a traditional simulator. These opportunities focus on associations between trains, such as connections between trunk and branch services, and terminal turnrounds of vehicles or crew.

Whilst a traditional simulation system, such as VST Comreco Rail Ltd.’s RailPlan, can recognise a simple association between an arriving trip and a subsequent departing trip at a terminus, delaying the departure in the event of late arrival, subject to a minimum turnaround time, the full range of possible associations seems to have been a step too far in design and modelling workload for a system which has its roots in engineering rather than train planning. Such associations include not just trips worked by the same set of vehicles, but multiple associations such as two trains combining to form one. Moreover,
associations between trains may derive from crew duties as well as vehicle workings, and therefore be subject to a quite different set of criteria.

**TrainPlan** supports **ResourcePlan**, an extension module aimed at vehicle and crew scheduling, and so has the potential to tap data on train associations from the planning system which creates them. It therefore becomes a relatively simple matter to examine the performance behaviour of timetables as it is affected by resourcing issues, for instance, in assessing the reliability of the same timetable but supported by different sets of vehicle and crew workings. Relevant issues for the resources plan potentially impacting upon reliability include the difference between actual and minimum turnaround times, and the extent to which vehicles and crews work over a number of routes in a working day.

10 Conclusions

Comprehensive, TPC-driven network simulators have a valuable role in studying specific issues within limited geographical areas, to support such functions as specification of signalling and evaluation of infrastructure options. In wider areas they are also vital in driving power supply simulations, where detail of each trains movement must be captured precisely.

However, when modelling complete train services over an extensive network, such a tool verges upon a liability, requiring trained and practised operators, incurring a severe workload to compile a model, with output potentially distorted by unnoticed input errors. The superficial precision of the required input data gives a false impression of validity. What is left out may be of more importance, but the task of gathering data to support additional input opportunities may be unrealistic.

I believe that if simulation of widespread networks is to be a task that makes good use of time and money, the challenge is to simplify the inputs rather than add to them. Accepting a coarser representation of train movements and the infrastructure bypasses much of the workload whilst preserving a valid representation of network effects.

References
