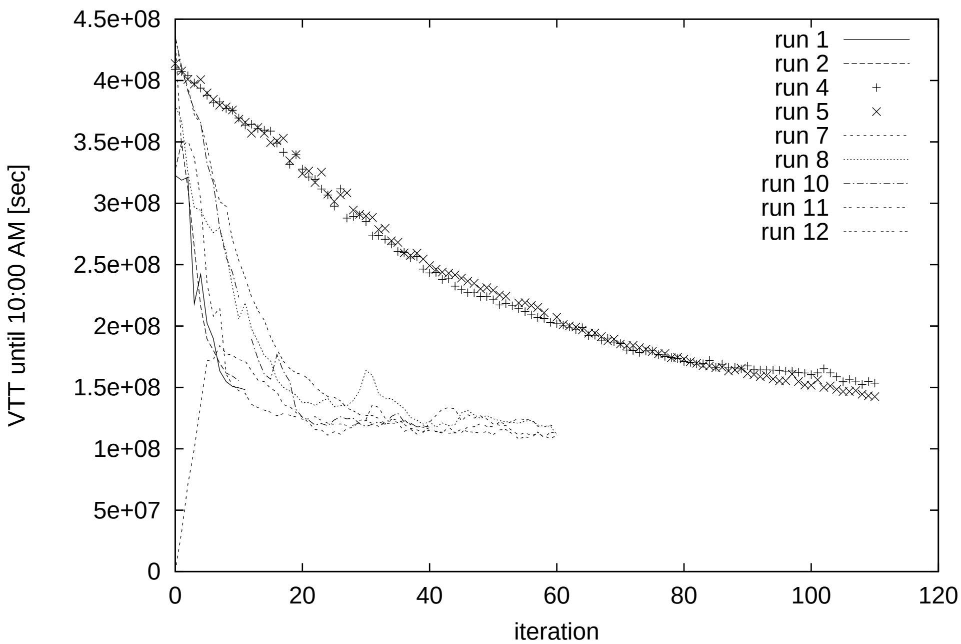
LA-UR 98-4601. This version: December 15, 1998. Draft, comments welcome
Approved for public release; distribution is unlimited

Issues of simulation-based route assignment

Marcus Rickert and Kai Nagel



Issues of simulation-based route assignment

K. Nagel and M. Rickert

Los Alamos National Laboratory, TSASA, Mail Stop M997, Los Alamos NM 87544, U.S.A., kai@lanl.gov, rickert@lanl.gov

Abstract

We use an iterative re-planning scheme with simulation feedback to generate a self-consistent route-set for a given street network and origin-destination matrix. The iteration process is defined by three parameters. We found that they have influence on the speed of the relaxation, but not necessarily on its final state.

1 Introduction

The traditional urban transportation planning process consists of four steps: trip generation, trip distribution, modal split, and assignment of routes onto the network. In recent years, more and more work has been invested to consider the dynamic aspects of these components. However, it has proven very difficult to define – let alone solve – the dynamic problem analytically (e.g. [1]).

Iterated microsimulations of traffic provide an alternative for the assignment portion. Several groups (e.g. [2, 3, 4, 5, 6, 7]) have used the iterative approach of *routing – microsimulation – feedback of travel-times* to obtain an assignment (route set) that is – within the accuracy permitted by any implicit or explicit stochasticity of the model – self-consistent. By self-consistency we mean that the assumptions about link travel times that are made during the routing period are consistent with the link travel times that are encountered in the microsimulation once all route plans are executed.

All these approaches are conceptually not very different from traditional assignment [8] except that the link delay function calculation is replaced by the microsimulation, which generates link delays from the simulated dynamics of the system. This has the advantage that the results are –at least in principle– dynamically more correct than the traditional link delay function, which cannot, for example, deal with queue build-up.

The currently existing simulation-based methods are simple relaxation methods. Note that traditional assignment also started out with simple relaxation methods, which became more sophisticated over time with the accumulation of knowledge. This indicates that over time researchers will probably be able to gain more insight into simulation-based assignment. The present paper discusses, based on simulation results for a Dallas scenario, some of the issues.

2 Context

The context of this paper is the so-called Dallas–Fort Worth case study of the TRANSIMS project [9]. The purpose of the case study was to show that a micro-simulation based approach to transportation planning such as promoted by TRANSIMS will allow analysis that is difficult or impossible with traditional assignment, such as measures of effectiveness (MOE) by sub-populations (stakeholder analysis), in a straightforward way. Many accompanying studies such as Refs. [10, 11, 12, 13, 14, 15] attempt to document the technology leading to and following up on the case study.

The underlying road network for the study (public transit was not considered) was a so-called focused network. It contained *all* links in a 5 miles times 5 miles study area, but got considerably "thinner" with further distance from the study area.[1] The full focused network contained 9864 nodes and 14751 mostly bi-directional links. The study area contained 2292 nodes and 6124 uni-directional links with a total length of 2276 lane kilometers – which is considerably larger than many networks used in other studies. A picture of the network can be found in Ref. [11], an excerpt is shown in Fig. 2.

The TRANSIMS design specifies to use demographic data as input and to generate, via synthetic households and synthetic activities, the transportation demand. The Dallas/Fort Worth case study was based on interim technology: parts of the demand generation modules were not yet available. For that reason, the study uses conventional 24-hour trip tables (production-attraction matrix, PA matrix) as starting point. The PA matrix was provided by the regional transportation planning authority (the North Central Council of Government, NCTCOG). The PA matrix roughly is a 24 hour origin-destination matrix, i.e. the metropolitan area of Dallas/Fort Worth is divided into approximately 800 zones (traffic analysis zones, TAZs), and the number of trips going from each zone to each other zone in a 24 hour period is given.

For the case study, the first thing that was done was to break down the PA matrix into individual trips [9]. For this, a time-of-day distribution according to land use in the destination zone was used. For example, traffic going to commercial zones mostly occurs in the morning. Also, starting and ending locations of trips were specified on the link level. The result was a table of approx. 10 million trips, all with a starting time, a starting location, and a destination location. From this table, all trips starting between 5am and 10am (ca. 3 million trips) were actually used.

Next, an "initial routing" step was done. All 3 million trips were routed according to "fastest path in an empty network" (i.e. using free speeds provided by the transportation authority). All trips that went through the study area in this step were retained, all other trips were removed. Note that this defines a base set of trips for all subsequent studies presented in this paper: All trips thrown out in this step can no longer influence the result of the studies. This base set contained approx. 300 000 trips. It was used to create three different initial plan-sets which will be discussed later on.

3 The micro-simulation

The above procedure does not only generate a base set of trips, but also an initial set of routes (called *initial plan-set*). These routes are then run through a traffic micro-simulation, where each individual route plan is executed subject to the constraints posed by the traffic system (e.g. signals) and by other vehicles. Note that this implies that the micro-simulation is capable of executing pre-computed routes – only very few micro-simulations currently have this capability although their number is growing. It also implies that, in the simulations, drivers do *not* have the capability of changing their routing on-line.[2]

We have used several micro-simulations for the Dallas study, ranging from extremely realistic (including, for example, lane changing for plan following, turn pockets, or pre-timed signal plans) [12] to a simple queue model [17]. The purpose of these studies was to test the robustness of the results against changes of

[1] Note that this "thinning out" of the network was not done in any systematic way and is explicitly *not* recommended. It was an ad-hoc solution because more data was not available, and because of limited computing capabilities.

[2] It is not that on-line re-routing is incompatible with the technology (see, e.g., [16]), but it has not generally been implemented and studied.

the micro-simulation. The results of these studies, in one case also in comparison with field measurements, are published in [15, 14].

In this paper, we are interested in the dynamics of the iterative procedure itself. For this purpose, we want to concentrate on a single micro-simulation, which has been used to systematically explore possible relaxation procedures. This micro-simulation, called PAMINA (PArallel MIcroscopic Network Algorithm), is based on the so-called cellular automata (CA) technique for traffic flow [18, 19, 20]. There is no necessity for this except the requirement of sufficient computational speed. PAMINA does not include signal plans, weaving and turn pockets, and lane changing for plan following. We expect this, though, to make less of a difference than one might think, due to the integration into the re-routing process. In particular, PAMINA is significantly more realistic than the link performance function used in traditional assignment. PAMINA uses the correct number of lanes, and since vehicles occupy space in a jam, this generates the "link storage capacity" for queue spill-back. It uses "averaged" signal plans, which generate, together with the number of lanes, the flow capacities. Thus, in most aspects, PAMINA is sufficiently realistic so that we can expect the technology to be representative for other micro-simulations that it could be replaced by. This intuition is backed up by our comparison studies [15, 14].

The reason to use PAMINA in spite of its somewhat restricted representation of reality is its superior computing speed. It ran more than 20 times faster then the TRANSIMS micro-simulation for this study, which was a combined effect of using faster hardware (the code is much easier to port to different hardware, thus being able to take advantage of new and faster hardware much sooner), less realism, and an implementation oriented towards computational speed. Further information on PAMINA can be found in [21, 22, 13, 16].

4 Feedback iterations and re-planning

The initial plan-set is obviously wrong during heavy traffic because drivers have not adjusted to the occurrence of congestion. In reality, drivers avoid heavily congested regions. We model that behavior by using *iterative re-planning* [23, 1, 24]: The micro-simulation executes the routes, which, as said above, have starting times between 5am and 10am. After this, the simulation is continued until noon to let jams dissolve. The travel times of all vehicles on each link are averaged into bins of duration Δt and recorded in a feedback data file. Throughout this study, we use $\Delta t = 15$ minutes.

After the termination of the simulation, the feedback data is used to re-plan a certain fraction f_r of all plans. The plans to be re-routed can be chosen randomly, or according to additional criteria. Together with all other "old" plans, they will constitute the route set for the next iteration.

After re-planning, the micro-simulation is run again on the new plan-set, more drivers are re-routed, etc., until the system is "relaxed", i.e. no further changes are observed from one iteration to the next except for fluctuations (all micro-simulations are stochastic). A typical iteration step takes about 6-8 minutes for pre-processing, 30-35 minutes for simulation using eight CPUs (250 MHz) on SUN Enterprise 4000, and 15-20 minutes for re-planning. This means that a single series of 60 iterations (as we ran many of) takes about 2 1/2 days on 8 CPUs of that machine.

5 Dynamical systems view

5.1 Consistency problem; equilibrium approach

Transportation-related decisions of people depend on what everybody else is doing. For example, decisions about mode choice, route choice, activity scheduling, etc., depend on congestion, caused by the aggregated behavior of others. From a conceptual viewpoint, this *consistency problem* causes a deadlock, since no agent can start the planning of her activities because she does not know what everybody else is doing.

In fact, this problem is well-known not only in transportation, but in socio-economic systems in general. The traditional answer is to assume that everybody has complete information and is fully "rational", i.e. that, for some given utility-function, each individual agent picks the solution that is best for herself. This means that each individual agent's decision-making process now is globally known, and so each individual agent can (in principle) compute everybody else's decision-making process conditioned on her own. Now, since everybody does exactly the same computation, one can replace the individual decision-making process by a global computation.

This is now a well-defined problem, which can in principle be solved. For example, in the typical well-structured 2-player one-shot problems often used in game theory [25], one can write down all possible behaviors of player 2 as a reaction of all possible moves of player 1. Since each player can assume that the other one will make the best possible move, one can compute what each player will do in any given situation. – In practice though, this turns out to be an extremely hard problem. For sufficiently complex games, the above exhaustive approach turns out to be computationally infeasible.

One is, though, not really interested in each players reaction to every possible move of the other player(s), but mostly in the question if there is a situation where each player has a certain strategy and no incentive to move away from this strategy. This is the traditional Nash equilibrium. In transportation, a typical example is the user equilibrium solution of the static assignment problem [26]: No driver (or traffic stream) can be better off by switching routes.

Traditional economics has often focused on the computation of such an equilibrium – because already that usually is an extremely hard problem. In traffic assignment, the Nash/user equilibrium solution turns out to be a "robust" solution in the following way: The Nash equilibrium solution is also the solution of an equivalent optimization problem, and under certain assumptions, there are no local minima besides the global minimum. In consequence, all algorithms that find a local minimum will find the same solution (in terms of the link flows). In other words: If one defines $C : x_i \to c_i$ as an operator that translates link flows into link costs, and $X : \{c_i\}_i \to \{x_i\}_i$ as an operator that translates link costs into a new assignment with the property that drivers switch to better routes if they existed in the last iteration, then the user equilibrium is a fixpoint (i.e. $X(\{C(x_i)\}_i) = \{x_i\}_i$) of the assignment dynamics; many mappings X will converge to that fixpoint; and there is only one basin of attraction.[3] This means that if one finds an algorithm that converges quickly, then the assignment problem is solved.

It is useful to recall again some of the assumptions that underly the traditional assignment approach:

[3]For formal treatments, this needs to be made more precise, see, e.g., [26]. For the purposes of this paper, which concentrates on intuition and on simulation results, we feel that the above version is sufficient.

- Agents are complete rational and they have complete information.
- It is possible to calculate link costs c_i from the flow assignments x_i only, without any further information.
- The only point of interest is the equilibrium point itself. – This point is important, since it means that any relaxation procedure to reach the equilibrium point is just a computational trick and does not have to be rooted in human behavior.

5.2 Adaptive systems approach

It is instructive to look at biological ecosystems for a minute. Here also, the behavior of everybody depends on everybody else. For example, an animal should not eat in an area where predators catch it. Yet, since we assume that animals are less capable than humans to perform organized planning and reasoning, nobody ever assumed that animals would pre-compute an optimal solution based on some utility function. Instead, one formulates the problem as one of *co-evolution*, where everybody's (mostly instinctive) behavior evolves in reaction to what is going on in the environment, constrained by the rules of genetical chemistry.

It is indeed this "eco-system" approach that more and more groups are also taking in transportation simulation and in the simulation of socio-economic problems in general. The advantage is that one does not have to make assumptions about properties of the system that are necessary in order to make the mathematics work. For example, one can just define rules how agents decide on switching routes, and let the simulation run. Clearly, for this no link delay function is necessary, since one can extract dynamically correct link delays directly from the micro-simulation.

5.3 Iterated transportation simulations as dynamical systems

It is instructive to see such iterated transportation simulations as dynamical systems rather than as assignment problems. These dynamical systems operate on two time-scales, which should not be confused:

- The simulationg generates a traffic dynamics that depends on the time-of-day. For example, a possible output of such a micro-simulation may be link delays as a function of when vehicles enter the link.
- The iterations generate "periods" of the game, often called "days" because, in some sense, one watches a day-to-day evolution of the co-evolution problem where everybody adapts their decision based on what everybody else does (or did in the past). This day-to-day evolution provides a mapping into itself. For example, route plans get executed by the micro-simulation, and based on this, a new set of route plans is generated.

It is the second process (the process of iterations) that we relate to in this paper.

Let us introduce some terminology. If one looks at the evolution of the system as a function of time-of-day, it can for example be described as a trajectory in $\sum_{i=1}^{N} f_i$ dimensions, where N is the number of particles (vehicles, travellers, "agents") and f_i is the number of their degrees of freedom (position, speed, ...) of the ith particle (Fig. 1). The trajectory of this system is thus an object with $1 + \sum_{i=1}^{N} f_i$ dimensions, where the additional dimension is the time-of-day dimension. If the iterations would reach a fixpoint, then the trajectory of the system at iteration $n + 1$ would exactly lie on top of the trajectory at iteration n. Thus, if one wants to include the dynamics of the transportation system in the description, one needs to look at mappings in this $(1+\sum_{i=1}^{N} f_i)$-dimensional space.

If one recalls static traffic assignment algorithms, they indeed do something similar: Route plans are mapped, via the link cost functions, into new route plans, and the iteration stops once the route plans do not change any more.

However, in iterated simulations we do not have the knowledge from the static assignment any more. So, the questions now are, for example: Does the iterative process converge, or does it, say, go into some kind of periodic or possibly chaotic attractor? If it converges, what does that mean? Can we show some uniqueness of the converged result, or are there multiple "basins of attraction"?

The problem gets confounded when the mappings are stochastic. Here, even if the dynamics is driven towards a fix point, the stochasticity will make the system jump around that fixpoint. At best, one can thus reach a time-independent density in phase space, i.e. for any given volume in phase space, the probability to be hit in the next iteration is constant in time. This would correspond to a stationary Markov process.

As long as that phase space density is Gaussian distributed, one could at least still use concepts like "the average behavior of the system". Unfortunately, there is no reason to believe that the outcomes will be Gaussian; for example, in our simulations we have found distinctively non-Gaussian fluctuations [11].

Then, we can also have a situation where the deterministic equivalent of the stochastic system is on a periodic or possibly chaotic attractor. In that case, at best one could hope to find time-independent phase space densities when averaging over enough iterations.

And last, we could have multiple basins of attraction of the iteration process, leading to solutions that depend on the initial conditions, on adaptation rules, and/or on random events.

6 The different studies

Clearly, there are many ways how one can execute the details of the iterative procedure. Optimally, some theory would be available in order to guide the selection of the most appropriate procedure for a problem at hand. Since we are currently short of a theory, we have instead performed many relaxation series with different set-ups in order to highlight some of the issues and in order to find reasonably fast methods for relaxation.

6.1 Different re-planner algorithms

We have used two different implementations of the re-planner. Let us call them RP1 and RP2. RP1 is the re-planner that was used for the Dallas/Fort Worth case study [9]. RP2 is a faster and less memory-consuming version that has been implemented since then. RP1 and RP2 are written according to the same specifications: they compute fastest paths based on 15-minute averages of link travel times using a time-dependent implementation of the Dijkstra algorithm [11, 27]. Time-dependence is accounted for in the following way: The micro-simulation reports the average link travel time of all vehicles *leaving* the link between, say, 8:00 and 8:15. RP2 then uses this link travel time for all Dijkstra calculations that *enter* the link during the same time period. RP1 uses this link travel time for all Dijkstra calculations that enter the link between 7:45 and 8:00 (thus "anticipating" congestion build-up). Clearly, both algorithms are somewhat sloppy here; newer implementations of our algorithm deal with this in a more precise way by actually calculating when, in the average, the vehicles had entered the link. Both RP1 and RP2 were tested and no significant differences were seen.

6.2 Different initial plansets

Although all iteration series use the same trip table, the initial plan-set could be re-computed. The plan-sets used were:

- Use free speed link travel times as link costs. Called FS.
- Use the geometrical lengths of links as link costs. Called SP for "shortest paths".
- Start with an empty route set in which all routes were initially de-activated. In this case, "replanning" a trip means: If the trip is touched for the first time, "activate" and route it. Once a trip is activated, it remains activated throughout the series. – The reason for this route set is to see if one can reach relaxation faster when loading the network incrementally with more traffic from iteration to iteration instead of all at once. Called VD for void.
- The initial plan-set used for the simulation of the TRANSIMS case-study (see [9]), called CAS.

6.3 Re-planning fraction

Choosing the re-planning fraction f_r is a trade-off between computational speed and stability of the iterative process. Large fractions allow fast re-planning of all routes in the initial route-set. As we will see, one prerequisite of relaxation is that most of the routes have been re-planned at least once. As the iteration proceeds, however, large fractions have the clear disadvantage of moving routes back and forth between similar alternatives (see, e.g., [11]).

6.4 Route selection

Having set a re-planning fraction f_r does not yet define *which subset* of the previous route set is to be rerouted. One can easily imagine a selection according to one or more criteria, i.e. (i) selecting routes traversing a certain area of interest (or congestion), (ii) characteristics of the driver, (iii) planned or actual travel-time, (iv) last re-planning of the route. In this paper, we concentrated on a small number of studies:

- Pick a fraction of routes randomly, independent of their age ("RND").
- Use scheduled re-planning fractions and pick randomly ("SCD"). More specifically, use the fractions (in this order) 3*20%, 3*10%, 3*5%, and 1*2% for run 2 and an additional 2% for run 1. See also Tab. 1.
- Use a probability linearly increasing with the age of a route ("AGE").

- The N routes are split into two groups: those routes that have never been re-planned yet, and those that have been re-planned at least once. For n iterations, we replan N/n plans of the first group, and an additional fraction of f_r of the second group. ("RDC", for "reduce" since the number of not re-planned routes is forcibly reduced).

RND and AGE eventually lead to a stationary age distribution $f(a)$ which can be analytically predicted. If $p(a)$ is the probability for a route of age a to be re-planned, the following equation has to hold:

$$f(a + \Delta a) = f(a) - p(a) f(a) \Delta a \ ,$$

where $\Delta a = 1$ for our case.

For $p(a) = f_r$, this can be solved exactly by

$$f_{rnd}(a) = f_r (1 - f_r)^a .$$

For a general solution, one can take the limit $\Delta a \to 0$ and then integrate the equation, leading to

$$f(a, p(...)) = e^{\int p(a) da} .$$

For example, for the linear age selection we use $p(a) = C_1\, a$ and $f(0) = f_r$, and we obtain a normal distribution

$$f_{lin}(a) = f_r e^{-1/2\, C_1\, a^2}$$

as an approximation to our discrete case. Note that the "tail" for large a is much thinner here, as desired. Fig. 6 depicts $f_{lin}(a)$ and $f_{rnd}(a)$, Fig. 7 depicts the resulting age distributions.

The age distribution of RDC is a mainly linearly decreasing function overlaid by a very shallow exponential decrease. "transims" yields a step-wise linear distribution also overlaid with a shallow exponential decrease. In both cases it does not make sense to speak of a stationary distribution.

Details about the iteration parameters can be found in Tabs. 1 and 2.

7 Results

7.1 Relaxation

Fig. 3 shows the number of vehicles in the study-area for run 4 with 110 iterations. After an initial phase in which routes are so poorly distributed that the whole systems grid-locks (see horizontal lines), it eventually improves and approaches a state with very small fluctuations.

7.2 Uniqueness

We first want to look at uniqueness (see Fig. 4).

As stated above, the question is if the relaxed answer may depend on the particular initial condition and the particular re-planning mechanism that was selected. Fig. 4 depicts the accumulated travel-time of all vehicles ("vehicle time traveled", VTT) plotted against the iteration number. All relaxation series seem to relax towards the same average VTT. Although this is certainly far from a mathematical or even computational proof, it is at least consistent with the assumption that there was only one large basin of attraction for the iterated simulations that we did.

This gets backed up by looking at the traffic patterns in the relaxed states (not shown) that always displayed similar structures. Furthermore, even when using different micro-simulations in the same re-planning structure, the traffic patterns came out comparable [15].

This is good news, or at least no indication of bad news. If iterated micro-simulations would turn out to have several basins of attraction, computational studies for analysis and forecasting would become much more cumbersome.

7.3 Stochastic fluctuations

Since our simulations are stochastic, they will result in different traffic patterns even when fed with exactly the same plan-set. In other words: Re-running any of the iterations with a different random seed will generate different traffic. In plots like Fig. 4, such fluctuations are visible on top of the changes that happen due to the re-planning. Clearly, even when the simulations relax towards the same average VTT, they display fluctuations from day to day even when no route re-planning takes place. This makes the application of "deterministic" relaxation criteria impossible, since with stochastic simulations one can at best obtain a constant probability density in phase space.

Sometimes, fluctuations can be fairly significant. In the TRANSIMS micro-simulation, one such example of this has been documented [28]. This indicates that the results of our simulations may display non-Gaussian distributions even of aggregated variables. Fig. 5 contains a schematic example. Having non-Gaussian distributions means that one needs to interpret averaged numbers with much more care.

7.4 The rate of relaxation

From a practical point of view, we are interested in reaching the relaxed state as quickly as possible. (This assumes that we consider the relaxed state as the solution of our question – see Sec. 9 (Discussion) for more remarks on this.)

Yet, from a theoretical point of view, the iteration procedure is actually best justified for small re-planning fractions. The reason is that we assume people to attempt to make improvements. But this only works if the traffic pattern encountered in the next iteration is similar to the last iteration, since this is where the link delay information comes from. As soon as this assumption does not hold any more, it is not clear where any of the iterated algorithms will go. In fact, in Fig. 4 (also see Figs. 9 and 8), runs 1 and 11 are examples for runs where the re-planning fraction may already be too high since they display significant jumps opposite the direction of relaxation which are larger than pure noise.

In Fig. 4, the algorithms clearly relax with different rates. Runs 4 and 5 are done with values of $f_r = 0.01$, i.e. only 1% of the whole plan-set is re-planned in each iteration. This fulfills the requirement of slow changes from iteration to iteration best, but it is also very slow. The main impediment to faster relaxation is, especially near the end, the decrease in plans that have *never* been re-planned is too slow. At the n-th iteration, the fraction of plans that have never been re-planned is

$$F_{never}(n) = (1 - f_r)^n \ .$$

At the 100th iteration, we have $F_{never} \approx 0.37$, i.e. more than one third of all trips is still routed according to the initial plan-set.

This observation led to the use of so-called age-dependent re-planning, that is, the probability of a plan to be selected for re-planning is a function of the number of iterations a since it was last re-planned. For example, run 11 was done with

$$p_{replan}(a) = C_1 \cdot a \ ,$$

where C_1 is a constant only dependent on the asymptotic re-planning fraction f_r:

$$C_1 = \frac{\pi f_r^2}{2} \exp\left(\frac{\pi f_r^2}{4}\right) .$$

In addition, since we want a high re-planning fraction for the first iterations, we set the initial age of all plans to 60. This results in a re-planning fraction of approximately 0.25 for the initial re-planning. After 10 iterations the effective fraction has almost decayed to the asymptotic fraction of $f_r = 0.05$ (see Fig. 6).

7.5 Fraction of trips that have been re-planned at least once

From the above, we note that the fraction of trips that have been re-planned at least once makes up a significant part of the relaxation progress. In order to monitor this effect more closely, Fig. 8 displays the VTT as function of the fraction that has been re-planned at least once. The curves of the different schemes end up closer together, but not so convincingly. To a certain extent, this is not unexpected: After a trip has been re-planned once, it also plays a role if it has been re-planned another time etc.

7.6 Accumulated re-planning fraction

As a compromise, Fig. 9 displays the VTT as a function of the accumulated re-planning fraction. This is simply the average number of times each trip has been re-planned. This plot clearly shows a much better data collapse. In this plot, one also sees that small re-planning fractions lead –in some sense– to a cleaner procedure: Series that use large re-planning fractions, most notably runs 1 and 11, display zigzagging trajectories.

8 Nested spatial resolutions

As described in Sec. 2, for the Dallas case study TRANSIMS used a spatially nested approach: (i) Trips were generated across the whole Dallas/Fort Worth region with its roughly 3 million inhabitants. (ii) A subset of these trips was generated by routing all these trips using free speeds but only retaining those that went through a five miles times five miles study area. (iii) This reduced trip table was the basis for the route assignment, yet the route re-planning was still done on the full network.

This nesting of the dynamical representation was done because of resource limitations: At the time of the study, we did not have the network data to simulate all of Dallas (the network for the planner, used in steps (ii) and (iii), did not have enough traffic flow capacity to run realistic simulations on it). For that same reason, nesting is done in many areas of quantitative science, and it also causes problems everywhere. Let us state a simple example to illustrate the problem: Assume we have a link (say, of a freeway), and we want to represent part of it by a precise micro-simulation, yet for pieces of out far away from the location of interest we want to use a less realistic model. Clearly, one of the two model representations will have a higher capacity. (It is practically impossible to match two simulations to *exactly* the same capacity.) Thus, depending on which model has the higher capacity, jams will form either to the entry of the study area, or at the exit. Such jams are completely spurious and artifacts of the method, yet they are clearly phenomena that a modeler has to deal with.

We will see a variant of exactly the same problem (the matching of capacities) later in this section. Before that, we want to turn to a simpler problem: That of loss of plans due to circumventing the study area.

8.1 Loss of plans and mean field correction

When one looks at the re-planning mechanism, what we actually do is to record delays inside the simulated study area, and to pretend that the rest of the network remains at free speeds. Since we compute routes on the complete planner network which includes major arterials and freeways outside the study area, this will have the effect that routes "around" the study area will look faster to the router than routes through the study area.[4] In consequence, the method will "push" plans out of the study area. The effect of this can be seen for runs 11 and 13 in Fig. 10, which shows the number of plans going through the study area. As long as traffic is not relaxed, congestion is strong, and more and more vehicles are indeed "pushed out" of the simulated area.

It is intuitively clear what one could do to prevent this effect: Measure the average delay caused by congestion inside the study area, and assume that the average delay outside the study area is the same ("mean field approximation"). This is indeed what was done with run 14 in which a correction factor was computed as average over the increase of travel-times of all links inside the study-area weighed by the lengths of the respective links. This factor was used to increase the travel-time on links outside the area accordingly. Fig. 11 shows the time-dependency of this factor which reaches values up to 400 for early iterations.

Yet, this lets us notice another nesting of our dynamics, which is this time on the temporal scale. Making *all* of the study area slow simply makes some trips so slow that they do not get to the study until very late.[5] In some sense, now trips do not get squeezed out spatially any more, but the system reacts by squeezing them out temporally. – Later in the iterations, when congestion is lower, most of them come back in.

Yet, the whole procedure has a side effect: Since overall traffic in the study area increases, the system is much closer to break-downs, which result in suddenly much higher travel times. This can be seen in Fig. 12, where the number of actually executed plans fluctuates much more for run 14 with congestion correction outside the study area than for run 11. – As a compromise, another approach was tried in run 15 for which we used the square-root of the correction factor and not the factor itself. This run still yields a higher number of executed plans than those without correction, yet does not generate any grid-lock anymore (see [16]).

[4] One could have reduced origins and destinations of the trip table to the study area, not leaving the router this flexibility. But this would have forced routes to certain entry and exit links into and out of the study area, thus also limiting the amount of possible adaptation to congestion.

[5] All trips that start before 10am were retained, no matter *where* they started, even when far away from the study area. Slow link speeds everywhere will make the planner assume that they get to the study area late, and will insert them at an accordingly late time.

8.2 Impedance mismatches ("queue feedback")

The next problem we want to look at is somewhat curious. It came to light when looking at traffic in the study area after 10am: Instead of slowly dwindling down, it remained constant at a low level for a long time. The reason for this were queues at the entrance points into the study area (see Fig. 13). Note that the effect of waiting there is not fed back into the planner, for the following reason:

- Even if this queue is long, its outflow is at best at capacity, which means that the first link inside the study area will be reported as "fast".
- The link where the queue theoretically sits on is outside the study area and thus not simulated and no delays are reported.

As stated in the introduction to this section, this is a typical albeit extreme case of "impedance mismatch". Links outside the study area do not have capacity constraints and so the planner can put arbitrarily many trips on it. And when these trips enter the study area, the study area links simply operate at capacity, which results in a reporting of (nearly) free speeds.

An intuitive way out is to add the waiting time just outside the study area to the link travel time for the first link in the study area. This is illustrated in Fig. 14. In this way, the router now recognizes that this is a slow route, and re-routes travelers through other options. Fig. 13 shows the result: no more pending vehicles towards the end of the simulation in run 13.

8.3 Remarks

It may seem that some of the artifacts of using a study area may have been preventable. Yet, there was a reason why we willingly accepted to "lose" trips through this mechanism. This reason was our intuition that, for the Dallas case study, we had slightly more trips than the network could handle. This intuition came from the simulation experiments, and it could have been due to an underestimation of capacity (i.e. a calibration problem with the micro-simulation), or due to an overestimation of demand, or due to both. Remarks from the Dallas MPO indicate that overestimation of demand may be part of the problem [29]. Thus, we deliberately accepted to have an *adaptive* mechanism that dealt with reducing demand, instead of having an external mechanism prescribed by the analyst.

At this point, it is important to note that a micro-simulation technology, such as presented here, is considerably more sensitive to overestimations of demand than traditional assignment models were. The reason is that, in a traditional assignment model, a link being above capacity just is a number, whereas in a dynamical micro-simulation it causes queues, which eventually spill back into other routes, and may cause grid-lock of the whole system. Thus, a precise matching between demand and network capacity will become one of the main challenge of micro-simulation based modeling, which indicates that the modeling of the demand needs to become part of the iterative process (see, e.g., [30, 31]).

9 Discussion

In spite of the above results, we are far away from a consistent understanding of the iterative micro-simulation procedure, especially for stochastic micro-simulations. In this section, we want to discuss some additional observations.

9.1 Behavioral justification

It is important to discuss the justification for the computational procedure.

First, traditional assignment assumes complete rationality for all travelers, i.e., each traveler chooses the or a fastest route. Most currently used simulation-based assignment methods do something similar, although when the simulations are stochastic most methods do not guarantee that drivers actually converge to routes with shortest expected travel times.

Second, all "relaxation" or "equilibrium" approaches assume that one is interested in the relaxed state, and it thus does not matter how the computation gets there. Yet, there is no reason to believe that actual transportation systems are in this relaxed state. It may very well be that real people stop switching routes once they have a "reasonable" solution, or that the system itself changes too often (for example because of construction sites).

This means that we need to develop methods to measure how close the real system is to relaxation (see, e.g., [10]), and we also need to think about the behavioral justification of our computational procedure. If the transients of the relaxation process turn out to be important, much more care needs to be applied to the modeling of the day-to-day decision-making of the travelers than when only the relaxed result matters.

Another problem comes from the fact that we use stochastic micro-simulations. With *deterministic* network loading procedures (e.g. traditional link travel time functions, deterministic micro-simulations (i.e. [3, 7]), other network loading procedures (i.e. [**?**, **?**]), one can obtain a clean definition of the equilibrium fix-point: If no traveler wants to switch routes, the situation is stable, because the iterations will display exactly the same behavior in the next iteration. However, for a stochastic simulation the same is no longer true. Even is a situation where nobody wants to switch routes based on the last iteration, the next iteration can be, due to stochasticity, different enough so that some travelers *now* want to switch routes. After some thinking, one recognizes that this implies that travelers may want to use *conditional strategies*, something like "if the freeway looks congested on the over-pass, I will stay on the arterial". And clearly, conditional strategies are not included in the current iterative re-planning framework. Two possibilities immediately come up: (i) Allow for conditional strategies in the routing part. (ii) Allow for on-line route changes during the simulation. While both approaches seem feasible, they have their drawbacks: Conditional strategies would necessitate much more intelligence on the side of the router than is currently implemented in most projects. And on-line re-routing, albeit possible to implement (i.e. [32]), is conceptually difficult since it needs a framework that is very different from the iterative re-planning framework.

9.2 Relaxation criteria

In Sec. 7, we mostly used the sum of all travel times (vehicle time traveled, VTT) as an indicator for relaxation; relaxation was defined as VTT no longer changing except for stochastic fluctuations from one iteration to the next. Although this is certainly valid as a criterion, it is also not very satisfying, since it is not related to any aggregated quantity that is directly related to the process. For example, it might be useful to plot the average improvement per re-routing event – relaxation would then be indicated by this quantity going towards zero in the average. Such tests need to be left to future work [33]; however, the results from [10] mentioned above give some information.

9.3 Nested dynamics, adaptive systems and artifacts

As explained in Sec. 8, using a reduced-size study area causes a plethora of problems. It is worth noticing that adaptive systems in general have a tendency to exploit artifacts of the representation rather than adapt according to the realistic part of the dynamics. Thus, the general recommendation has to be that, in transportation science, we should avoid nested representations of the dynamics as much as possible until we have a better understanding of the modules and their interactions.

10 Conclusion

Transportation micro-simulations are replacing the traditional link delay function approach to traffic assignment. This makes it possible to represent the dynamics considerably more truthful to reality. Yet, the scientific foundation of the assignment process needs to be built again with these new methods. Systematic simulation experiments are a powerful tool here since they allow systematic exploration of the space of the possible dynamics yet is less restricted towards realism than traditional mathematics.

For a Dallas scenario coming from a realistic case study, we have performed many series of the assignment process, always using the same fast but reasonably realistic micro-simulation, but varying different parameters that define the assignment process. We showed that there is no indication that different paths towards relaxation can yield different results, which is good news. However, in most other aspects, using stochastic micro-simulations instead of link delay functions in the network loading is currently considerably more challenging, for example in terms of necessary computer time and because of non-Gaussian fluctuations. Nevertheless, we presented computational evidence that the trick to fast relaxation in a first order method is to allow the whole population to adjust routes at least twice, and to do this in as few actual iterations as possible. However, if one attempts to reach this in too few iterations, i.e. by allowing too many people to readjust routes from one iteration to another, the traffic situation in the last iteration is no longer representative for the traffic situation in the next iteration, and one gets oscillations instead of relaxation.

The Dallas scenario was run with a nested multi-scale approach, with a simulated five miles times five miles study area that was actually micro-simulated, and a much larger "focused" network that was used for routing. We highlighted problems and solutions of this approach, which indicates that a considerably better understanding of the procedure will be necessary before such a multi-scale approach can be reliably used for decision-making. Yet, a micro-simulation based approach to traffic assignment even with much larger systems would be possible with our software; the main restriction for Dallas was the availability of the data, not the computation. Clearly, the technology for the extended use of micro-simulation in transportation is there, and the work to understand the dynamics of these large interacting systems has just started.

Acknowledgements

We thank A. Bachem, R. Schrader, C. Barrett, and R.J. Beckman for supporting MR's and KN's work as part of the traffic simulation efforts in Cologne ("Forschungsverbund Verkehr und Umwelteinwirkungen NRW" [34]) and Los Alamos (TRANSIMS). We also thank them, plus P. Wagner, Ch. Gawron, and S. Krauß for help and discussions. We thank the TRANSIMS research and software teams for discussions and the processing of the data which served as input for the micro simulation. North Central Texas Council of Government (NCTCOG) provided road network and origin-destination data on which the simulations were based.

The work of MR was supported in part by the "Graduiertenkolleg Scientific Computing Köln".

This work was performed in part at Los Alamos National Laboratory, which is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36.

References

[1] David E. Kaufman, Robert L. Smith, and Karl E. Wunderlich. User-equilibrium properties of fixed points in iterative dynamic routing/assignment methods. *Transportation Research C*, forthcoming.

[2] R.H.M. Emmerink, K.W. Axhausen, P. Nijkamp, and P. Rietveld. Effects of information in road transport networks with recurrent congestion. *Transportation*, 22:21, 1995.

[3] H.S. Mahmassani, T. Hu, and R. Jayakrishnan. Dynamic traffic assignment and simulation for advanced network informatics (DYNASMART). In N.H. Gartner and G. Improta, editors, *Urban traffic networks: Dynamic flow modeling and control*. Springer, Berlin/New York, 1995.

[4] K. Nagel. Individual adaption in a path-based simulation of the freeway network of Northrhine-Westfalia. *International Journal of Modern Physics C*, 7(6):883, 1996.

[5] C. Gawron. An iterative algorithm to determine the dynamic user equilibrium in a traffic simulation model. *International Journal of Modern Physics C*, 9(3):393–407, 1998.

[6] M. Van Aerde, B. Hellinga, M. Baker, and H. Rakha. INTEGRATION: An overview of traffic simulation features. 1996. A paper accepted for presentation at the 1996 Transportation Research Board Annual meeting.

[7] DYNAMIT, see http://its.mit.edu.

[8] Y. Sheffi. *Urban transportation networks: Equilibrium analysis with mathematical programming methods*. Prentice-Hall, Englewood Cliffs, NJ, USA, 1985.

[9] R.J. Beckman et al. TRANSIMS–Release 1.0 – The Dallas-Fort Worth case study. Los Alamos Unclassified Report (LA-UR) 97-4502, 1997.

[10] T. Kelly and K. Nagel. Relaxation criteria for iterated traffic simulations. *International Journal of Modern Physics C*, 9(1):113–132, 1998.

[11] K. Nagel and C.L. Barrett. Using microsimulation feedback for trip adaptation for realistic traffic in Dallas. *International Journal of Modern Physics C*, 8(3):505–526, 1997.

[12] K. Nagel, P. Stretz, M. Pieck, S. Leckey, R. Donnelly, and C.L. Barrett. TRANSIMS traffic flow characteristics. Los Alamos Unclassified Report (LA-UR) 97-3530, 1997. Earlier version: Transportation Research Board (TRB) preprint 981332.

[13] M. Rickert and K. Nagel. Experiences with a simplified microsimulation for the Dallas/Fort Worth area. *International Journal of Modern Physics C*, 8(3):483–504, 1997.

[14] K. Nagel, M. Pieck, P.M. Simon, and M. Rickert. Comparison between three different micro-simulations and reality in Dallas. Los Alamos Unclassified Report (LA-UR) 98-2944, 1998.

[15] K. Nagel, M. Rickert, and P.M. Simon. The dynamics of iterated transportation simulations. Los Alamos Unclassified Report (LA-UR) 98-2168, 1998. Presented at the 3rd TRIannual Symposium on Transportation ANalysis (TRISTAN-III) in San Juan, Puerto Rico.

[16] M. Rickert. *Traffic simulation on distributed memory computers.* PhD thesis, University of Cologne, Cologne, Germany, 1998.

[17] Patrice Simon and Kai Nagel. Simple queueing model applied to the city of Portland. Los Alamos Unclassified Report (LA-UR) 98-9303, 1998. Also Transportation Research Board 1999 Annual Meeting Paper No. 99 12 49.

[18] K. Nagel. Freeway traffic, cellular automata, and some (self-organizing) criticality. In R.A. de Groot and J. Nadrchal, editors, *Physics Computing '92*, page 419, Prague, 1992. World Scientific.

[19] K. Nagel and M. Schreckenberg. A cellular automaton model for freeway traffic. *Journal de Physique I France*, 2:2221, 1992.

[20] K. Nagel, D.E. Wolf, P. Wagner, and P.M. Simon. Two-lane traffic rules for cellular automata: A systematic approach. *Physical Review E*, 58(2):1425–1437, 1998.

[21] M. Rickert and P. Wagner. Parallel real-time implementation of large-scale, route-plan-driven traffic simulation. *International Journal of Modern Physics C*, 7(2):133–153, 1996.

[22] M. Rickert, P. Wagner, and Ch. Gawron. Real-time traffic simulation of the German Autobahn Network. In F. Hoßfeld, E. Maehle, and E.W. Mayer, editors, *Proceedings of the 4th PASA Workshop.* World Scientific, Singapore, 1996.

[23] Alfredo Garcia and Robert L. Smith. On the convergence of iterative routing-assignment procedures in dynamic traffic networks. Technical Report 95-26, University of Michigan Department of Industrial and Operations Engineering, Ann Arbor MI 48109, November 1995.

[24] J. Bottom, M. Ben-Akiva, M. Bierlaire, and I. Chabini. Generation of consistent anticipatory route guidance. Presented at the 3rd TRIannual Symposium on Transportation ANalysis (TRISTAN-III) in San Juan, Puerto Rico, 1998.

[25] K. Binmore. *Fun and Games – A text on game theory.* D.C. Heath and Company, Lexington, MA, U.S.A., 1992.

[26] Y. Sheffi. *Urban transportation networks: Equilibrium analysis with mathematical programming methods.* Prentice-Hall, Englewood Cliffs, NJ, USA, 1985.

[27] R.R. Jacob, K. Nagel, and M.V. Marathe. A computational study of routing algorithms for realistic transportation networks. Los Alamos Unclassified Report (LA-UR) 98-2249, 1998.

[28] K. Nagel. Experiences with iterated traffic microsimulations in Dallas. In D.E. Wolf and M. Schreckenberg, editors, *Traffic and granular flow II*, Heidelberg, 1998. Springer. In press, also Los Alamos Unclassified Report (LA-UR) 97-4776, see www-transims.tsasa.lanl.gov/research_team/.

[29] K Cervenka. Personal communication.

[30] P. Wagner and K. Nagel. Microscopic modeling of travel demand: Approaching the home-to-work problem. Los Alamos Unclassified Report (LA-UR) 98-3368, 1998. Earlier version also Transportation Research Board 1999 Annual Meeting Paper No. 99 09 19.

[31] Jörg Esser and Kai Nagel. In preparation.

[32] M. Rickert. *Traffic simulation on distributed memory computers.* PhD thesis, University of Cologne, Cologne, Germany, 1998.

[33] J. Esser et al. In preparation.

[34] Forschungsverbund für Verkehr und Umwelt (FVU) NRW. See http://www.zpr.uni-koeln.de/Forschungsverbund-Verkehr-NRW/.

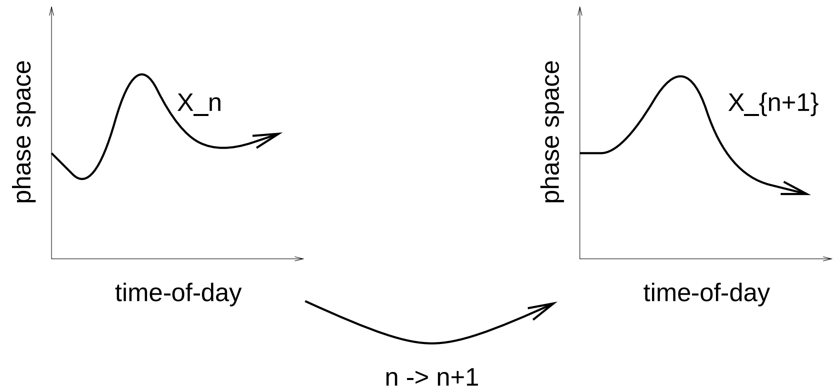


Figure 1: Schematic representation of the mapping generated by the feedback iterations. Traffic evolution as a function of time-of-day can be represented as a trajectory in a high dimensional phase space. Iterations can be seen as mappings of this trajectory into a new one.

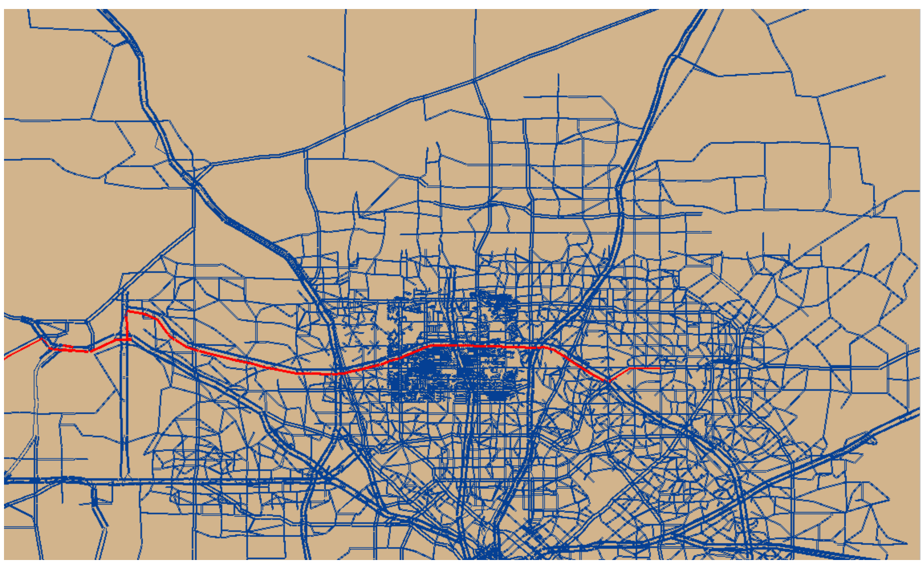


Figure 2: Partial view of the planning area. The simulation area can be recognized by the denser network at the center of the picture.

run	1	2	4+5	7	8	10	11	12
init. route-set	CAS	CAS	FS	VD	SP	FS	FS	FS
iterations	11	10	110	60	60	20 + 20	60	60
fraction f_r	SCD	SCD	1%	5%	5%	abs.5%+rel.1%	5%	5%
selection	RND	RND	RND	AGE	AGE	RDC	AGE	RND

Table 1: Parameter combinations of iteration runs. Note that for runs 1 and 2 the re-planning fractions were individually given ("scheduled") for each iteration.

run	13/16	14	15	17
route-set	FS	FS	FS	FS
iterations	60	80	80	60
fraction f_r	0.05	0.05	0.05	0.05
selection	AGE	AGE	AGE	RND
mean-field correction	-	lin.	sqrt	-
queue feedback	yes	-	-	yes

Table 2:

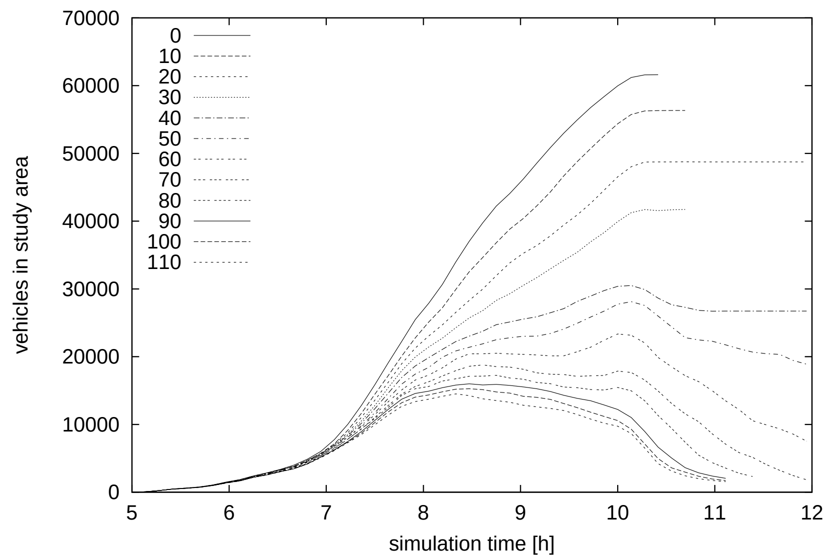


Figure 3: Number of vehicles in the study-area for run 4

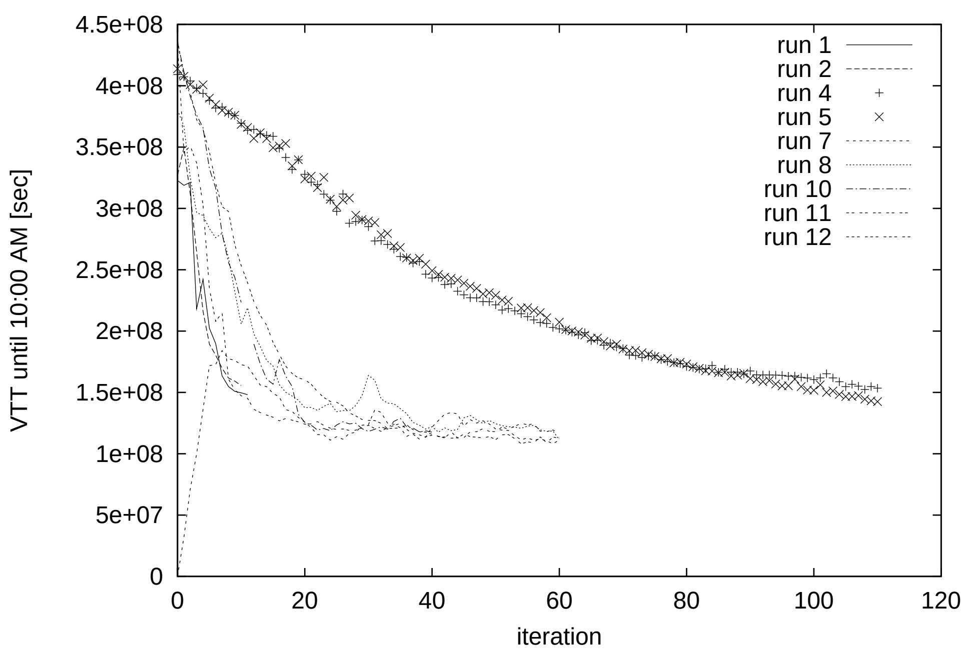


Figure 4: Relaxation of the iteration process.

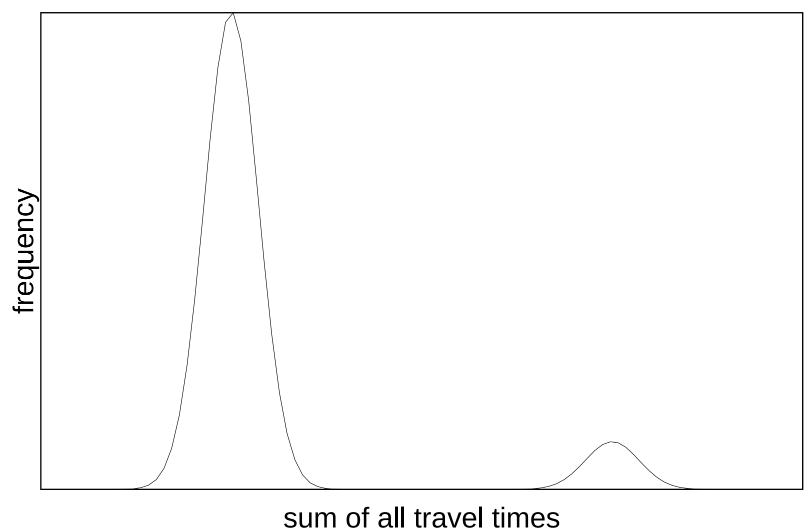


Figure 5: Schematic distribution of VTT (vehicle time traveled).

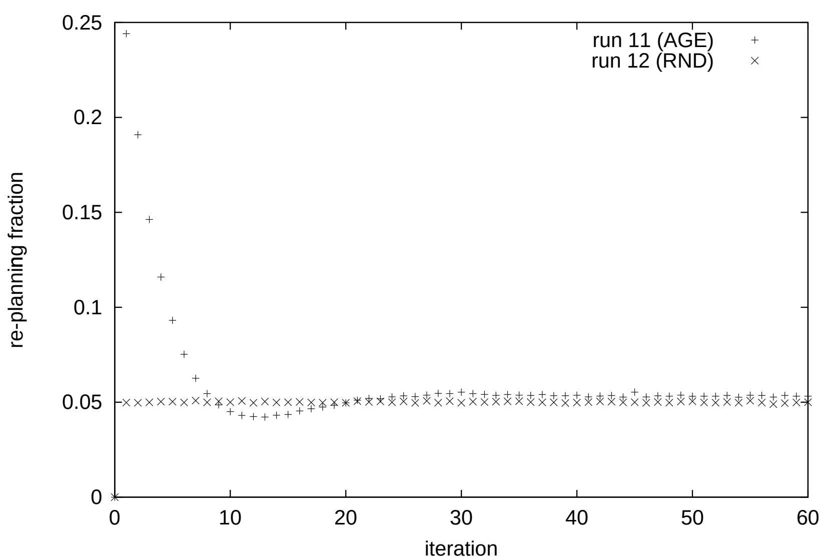


Figure 6: Effective re-planning fraction for the selection schemes RND and AGE.

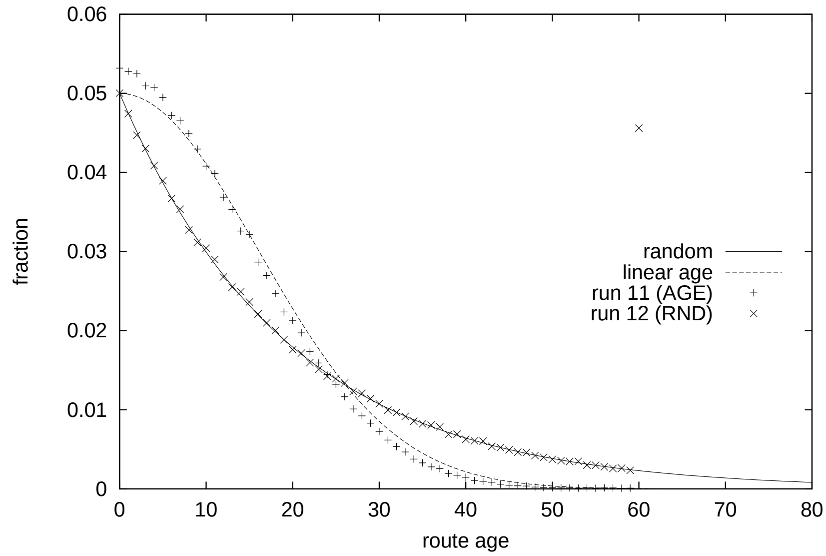


Figure 7: Age distributions for selection schemes RND and AGE after 60 iterations. Note that even with a re-planning fraction of $f_r = 0.05$ for run 11, there is still a fraction of 4.5% (see isolated point) that has been re-planned.

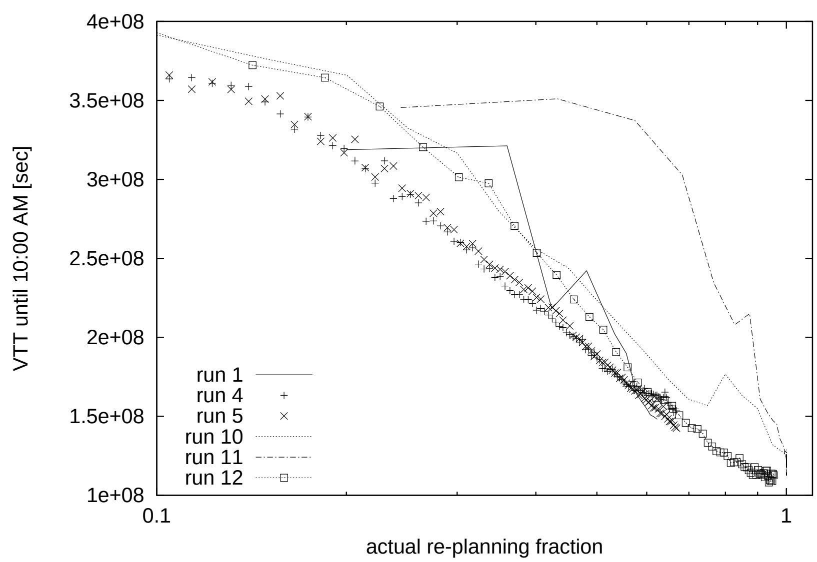


Figure 8: VTT vs. the actual re-planning fraction, i.e. the fraction of routes that have been re-planned at least once.

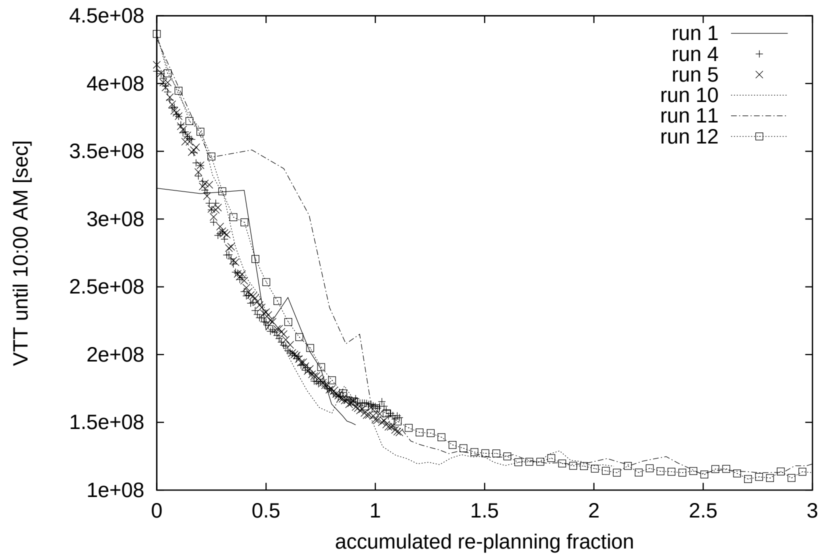


Figure 9: VTT vs. accumulated re-planning fraction: most curves collapse into one.

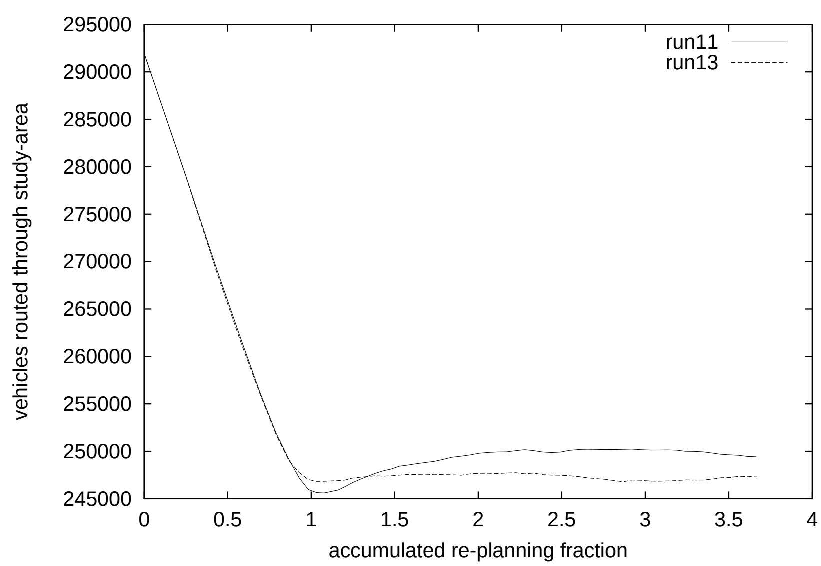


Figure 10: Vehicles routed through study-area: since the travel-time on links outside the study-area are not influenced by the feedback, the router starts to move routes outside the study-area.

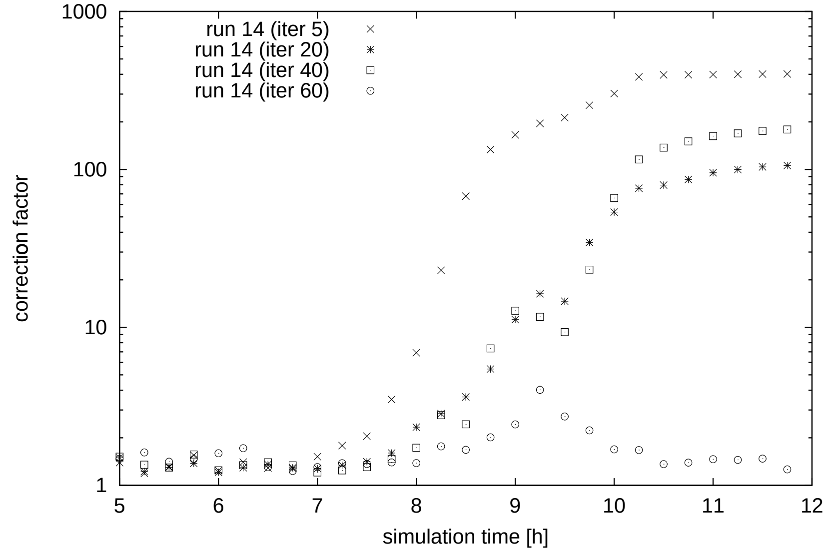


Figure 11: Mean-field feedback: the correction factor shows by what factor the travel-times inside the study-area have increased with respect to the free-flow case.

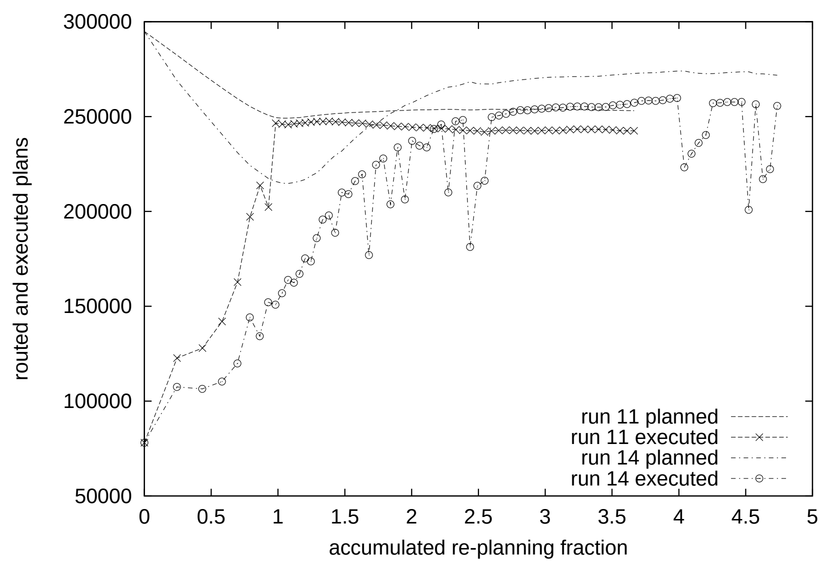


Figure 12: Mean-field feedback: routed and executed plans. In run 14 the correction factor forces so many vehicles back into the study-area that the simulation grid-locks from time to time. Note the initial dip of routed vehicles in run 14 due to the large number of vehicles outside the time-window (= entering the simulation area after noon).

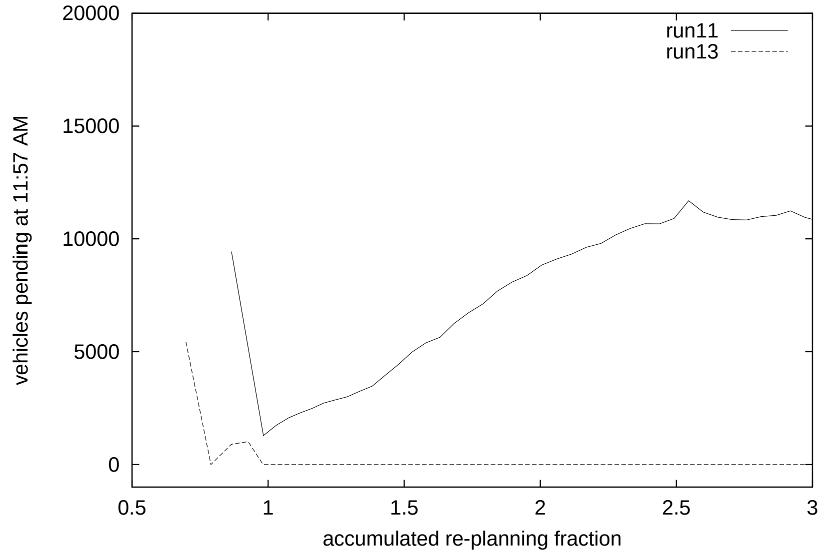


Figure 13: Vehicles pending at 11:57AM: insufficient feedback for feeding links into the study-area results in queue build-up and thus in a continuous flow of "late-comers". In run 13 the waiting time in queues is fed back which eliminates pending vehicles at the end of the simulation.

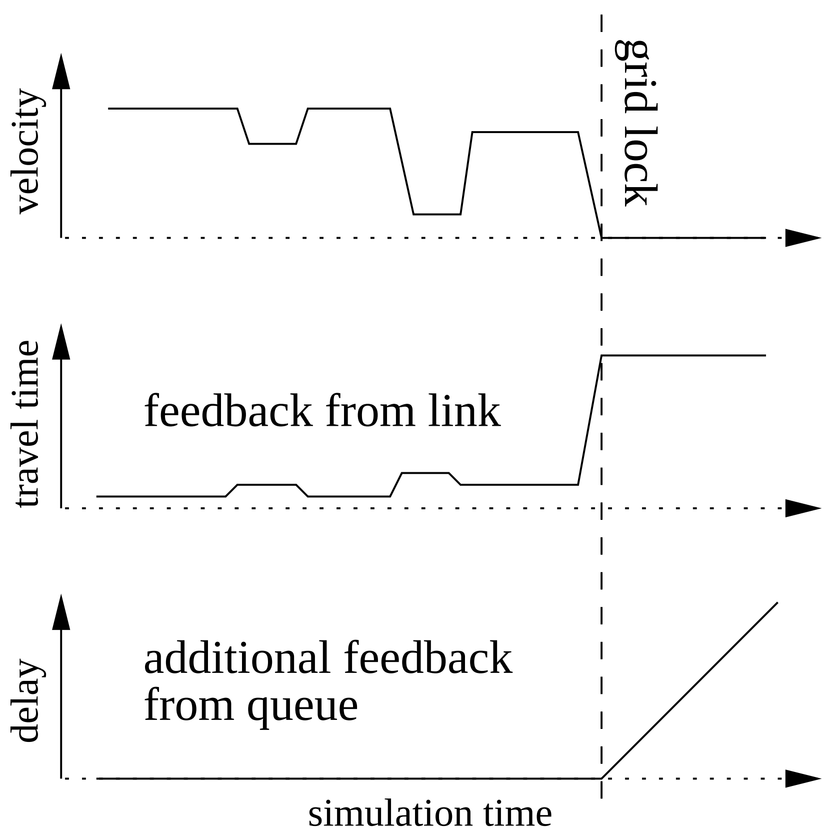


Figure 14: Queue feedback: in addition to the travel-time spent on links in the study-area itself, the queue feedback also considers the time spent in a queue of a feeding link.