Los Alamos Unclassified Report LA-UR 99-10 (This version: January 25, 1999)

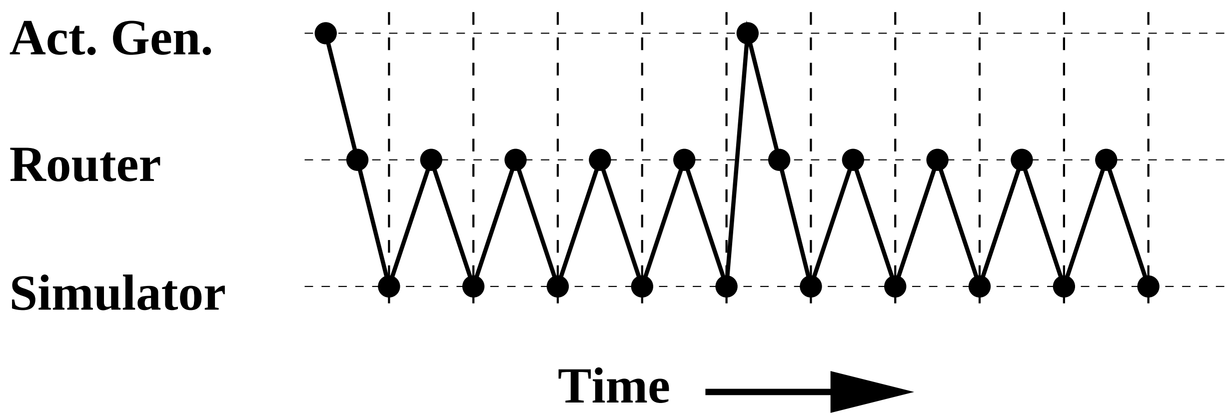
Approved for public release; distribution is unlimited

Census-based travel demand generation for transportation simulations

Jörg Esser and Kai Nagel

~~To appear in: Traffic and Mobility, edited by M. Schreckenberg et al, Springer, 1999~~

Published in: Census-based travel demand generation for transportation simulations. In: W. Brilon, F. Huber, M. Schreckenberg, H. Wallentowitz, Traffic and Mobility: Simulation - Economics - Environment. Springer, 1998, pages 135-148.



Published in: Census-based travel demand generation for transportation simulations.
In: W. Brilon, F. Huber, M. Schreckenberg, H. Wallentowitz, Traffic and Mobility:
Simulation - Economics - Environment. Springer, 1998, pages 135-148.

Census-based travel demand generation for transportation simulations

Jörg Esser and Kai Nagel

Los Alamos National Laboratory, Mail Stop M997,
Los Alamos NM 87545, U.S.A.

January 25, 1999

Abstract

During the last decades, a lot of progress has been made in understanding the dynamics of traffic flow models. Real world applications of these models require the ability to model traffic in networks, which has been an important research topic lately. We focus on the problem of travel demand generation in transportation networks. Our investigations are based on real world data for the Portland/Oregon area. We present a microscopic approach for iterative activity assignment exemplarily for home-to-work trips. It provides a method to generate real-world macroscopic data – in our case it is the travel time distribution resulting from census data – in a network traffic simulation under simulation feedback. The underlying assignment is based on a simple ansatz to split the probability of choosing a workplace in a particular distance into a term which describes the accessibility of workplaces, and the individuals' function of travel time acceptance. In combination with the census data, this approach provides the macroscopic acceptance function, which turns out to be an exponentially decaying function plus a 'repulsive' behavior for small travel times. Furthermore, these investigations demonstrate that iterative activity assignment on a microscopic level is computationally feasible even for realistically sized transportation systems.

Introduction

Traffic flows in transportation systems are driven by travel demands, which in turn result from households' needs and/or desires of participating in activities out of home characterized by activity type (e.g. working, shopping, sleeping), activity location, and activity time. These patterns are based on individual decision processes, where individuals take into consideration which traffic conditions they encountered in the past. For example, they adjust their shopping time and/or location according to traffic jams which they experienced the last trips.

There are fundamentally different approaches to reproduce realistic traffic in a laboratory system. For example, there are online-simulations, where traffic counts are tuned in the simulation in real time (e.g. [1]). They can be used as data source for intelligent transportation systems, for example for online routing systems or for dynamic traffic light control. A completely different approach is to look at long term impacts of travel demands resulting from given land use data by generating individual activity schedules and routes. In this context, the interactions between individual decision processes (regarding activity and route choice) and resulting traffic flows need to be considered. Without this, it will not be possible to represent realistic travel demands regardless of the underlying transportation simulation approach [2, 3, 4].

In this paper, we want to follow Ref. [5] and restrict ourselves to traffic from home to work. We assume that we know where all working people in a city live, and where all workplaces are. The problem is thus to match homes with workplaces in a realistic way.

An example of such a solution is the classic "Hitchcock" solution [6], where the workplace assignment is done in such a way that the overall sum of all travel times is minimized. This clearly results in much shorter trips than in reality.

The traditional approaches that are closest to what we want to present here are the so-called discrete choice models [7]. In this approach, the utility V_i of an alternative i is assumed to have a systematic component U_i and a random component ϵ_i. Under certain assumptions for the random component this implies that the probability p_i to select i is

$$p_i = \frac{e^{\beta U_i}}{\sum_i e^{\beta U_i}} \ . \tag{1}$$

p_i could for example represent the probability to accept a workplace that is i seconds away. If i is indeed taken as time, U_i is negative, and it follows an S-shaped curve, being flat for low i, steeper for medium i, and flat again for high i [8]. By this approach, our above location choice problem would be solved by weighting each given workplace in time-distance i by p_i and then making a random draw in these probabilities.

Clearly, for the discrete choice approach one needs to know the function βU_i. Also, once having obtained this function from, say, a survey, there is still no guarantee that a simulation based on this will generate realistic travel time distributions. In this paper, we thus want to present an approach where the "psychological" function βU_i can be obtained from "observed" travel time distributions, using new methods of micro-simulating large geographical regions. The kernel idea is that an observed travel time distribution $p_{obs}(t)$ can be decomposed into an accessibility part $p_{access}(t)$ and an acceptance function $p_{choice}(t)$ (traditionally called choice function):

$$p_{obs}(t) = p_{access}(t) \times p_{choice}(t) \ . \tag{2}$$

$p_{access}(t)$ is the probability that there is a workplace at time-distance t; $p_{choice}(t)$ is the probability that a prospective worker will accept this travel time. This decomposition also allows to separate the network specific distribution $p_{access}(t)$ from the travel time acceptance function, which is supposed to be of more general nature.

Given a micro-simulation of traffic, $p_{access}(t)$ can be derived from the simulation result. For a given home location (and a given assumed starting time), one can build a tree of time-dependent shortest paths, and every time one encounters a workplace, one adds that to the count for travel time t. The challenge is that this result depends on the traffic: Given the same *geographic* distribution of workplaces, these are farther away in terms of travel time when the network is congested than when it is empty. That is, given $p_{choice}(t)$, we can obtain $p_{access}(t)$ via micro-simulation, i.e. $p_{access}(t) = G(p_{choice}(t))$, where G is the micro-simulation. The problem then is to find the macroscopic (i.e., averaged over all trips) function $p_{choice}(t)$ self-consistently such that

$$p_{obs}(t) = G(p_{choice}(t)) \times p_{choice}(t). \tag{3}$$

Approach

The approach that we use is a regular relaxation technique. We start with a guess for $p_{choice}(t)$ and from there generate $G(p_{choice}(t))$ via simulation. A new guess for $p_{choice}(t)$ is obtained via

$$p^{n+1}_{choice}(t) = p_{obs}(t)/G(p^{n}_{choice}(t)) \ . \tag{4}$$

A fraction p_{act} of all travelers will do their workplace selection again, using the new p_{choice}. G is generated again via micro-simulation, and this is done over and over again until a sufficiently self-consistent solution for $p_{choice}(t)$ is found.

We use real census data for $p_{obs}(t)$ (Fig. 1). People usually give their travel times in minute-bins as the highest resolution. Since our simulation is driven by one-second time steps we need to smooth the data in order to get a continuous function instead of the minute-histogram. Many possibilities for smoothing exist; one of them is the beta-distribution approach in Ref. [5]. We encountered problems with that particular fit in our approach for small travel times: Since that fit grows out of zero very quickly, the division p_{obs}/p_{access} had a tendency to result in unrealistically large values for very small travel times. We therefore used a piecewise linear fit with the following properties: (i) For travel time zero, it starts at zero. (ii) At travel times 2.5 min, 7.5 min, 12.5 min, etc. every five minutes, the area under the fitted function corresponds to the number of trips shorter than this time according to the census data.

Obtaining $G(p_{choice}(t))$ itself is by no means trivial. It is now possible to micro-simulate large metropolitan regions in faster than real time, where "micro"-simulation means that each traveler is represented individually. We use a simplified queuing type traffic flow model described in [9]. However, even if one new the true origins (home locations) and destinations (workplaces), one would still need to find the routes that each individual takes. This "route assignment" is typically done via another iterative relaxation, where, with location choice fixed, each individual attempts to find faster routes to work. At this point we refer to [10, 11] for detailed information about the route-relaxation procedure; see also Fig. 2 and its explanation later in the text.

Once $p^{n+1}_{choice}(t) = p_{obs}(t)/G(p^{n}_{choice}(t))$ is given, the workplace assignment procedure works as follows: The workers are assigned in random order. For each employee the distances t for all possible household/workplace pairs *[hw]* are calculated, while the home location h is fixed and taken directly from the household data for each employee. Let t_{hw} be the resulting travel time for one particular *[hw]* and $n_{wo}(w)$ the number of working opportunities at workplace w. Then, an employee in household h is assigned to a working opportunity at place w with probability

$$p_{hw} \sim n_{wo}(w) p_{choice}(t_{hw}). \tag{5}$$

In addition to work location, home-to-work activity information also includes the times when employees start their trip to work. These are also directly taken from the household data.

For the iteration runs presented in the next section we have used a parallel implementation of the router module and a non-parallel implementation of the microscopic traffic simulator and the activity generator. Running this on a 250 MHz SUN UltraSparc architecture takes less than one hour computational time for one complete iteration run including activity generation, route planning and running the traffic simulator.

Iterative feedback experiment

Study area

Our investigations were are part of ongoing research efforts within the TRANSIMS (TRansportation ANalysis and SIMulation System) project at the Los Alamos National Laboratory. A reduced street network of Portland with 8,564 nodes and 20,024 links, where each link represents one driving direction between two nodes, serves as testing field. For this area a synthetic population of 1,415,900 individuals was generated based on a census from 1990 using an algorithm described in [12]. The resulting household data contains very detailed information (e.g. number of persons, employees, children and cars per household). As mentioned in the introduction, we focus on home to work activities. Thus, only people who work out from home are considered in the following; these account for about $520,000$ individuals in the synthetic population. Following the purpose of having a minimum model we neglect more detailed individual information (e.g. age, employees' salaries); we do not distinguish different types of employees. In addition to the household data, we use detailed land use data to extract locations and sizes (i.e. number of employees) of companies. For our simulations we make the assumption that all employees work within the study area, since there is no land use and household data available for the surrounding areas. This also means that workplaces can only be found up to a maximum distance. We will refer to this *finite size effect* later when we discuss the workplace distribution.

Simulation results

Fig. 2 shows how we run the iterative feedback in our experiment: Starting out with a fixed synthetic population, we assign people to workplaces, generate the routes as shortest paths, and run the simulation with these routes from 4am till 12pm. After this, we reroute people four times, while home-work relations are kept constant. Then, workplaces are re-assigned and so on. During the workplace re-assignment, each individual is re-assigned with probability $p_{ra} = 0.3$ to avoid over-reactions.

To understand the dynamics of the feedback on a macroscopic scale we first look at the overall travel time (i.e. the sum of all individuals travel times). Fig. 3 shows the total travel time versus the simulation run index (i.e. a sequence of iterative re-assignment according to Fig. 2).

In each workplace assignment step, worker are assigned to workplaces in a way that their expected travel times match the census travel time distribution. Roughly spoken, this means all individuals try do drive an average travel time, which remains constant for all iteration steps, while they face different traffic flow patterns in the network (represented by link travel times). For example, the initial workplace assignment generates too much travel demand, since it is based on the assumption of free speed link travel times. In other words, since initially each driver assumes that the network is empty, the initial workplace assignment generates trips with long geographical distances in order to reproduce the census data. The result of this is a lot of congestion occurring in the first simulation runs, so that in the first workplace re-assignment, individuals are assigned to shorter spatial home-to-work distances to keep the same average travel time. Once individuals are re-assigned, the origin-destination relation for the re-assigned trips and, by this, for the travel demand structure in the network is changed. It takes again some route re-planning steps to adjust routes to the new travel demand structure. In Figs. 4 and 5, snapshots of the simulation are shown for the first and the final simulation run, respectively. The links are colored according to the quotient of average speed and free flow speed; red represents low average speed values. These snapshots show that the overall travel demand is indeed decreased by the re-assignment procedure.

The dynamics of this iterative feedback loop is driven by the probability to find a working opportunity in a given travel time distance. For that reason, it is worth to have a look at the workplace distribution $p_{access}(t)$. Fig. 6 (Top) shows this distribution exemplarily for different activity iteration steps. The initial distribution based on the empty free speed network is linear for travel times up to about $900sec$. This can be easily understood by considering that the circumference of a circle is proportional to the radius and the workplaces are pretty much homogeneously distributed. For larger travel times, the chance to find workplaces decreases because of the finite size of the study are.

The distributions for later iterations also start out linearly but with a smaller slope, because the average speed on the links is lower than free speed. In addition, the slope fluctuates considerably for those assignments that are based on simulation feedback (i.e. all except the initial assignment). This is due to inhomogeneities with regard to travel times due to capacity constraints in the street network. For example, consider an individual that lives in an area that is connected to the rest of the network only via streets where capacities are exceeded, while within this area no congestion occurs. For this individual, the distribution $p_{access}(t)$ increases steeply for travel time distances within this area. However, once he/she tries to reach workplaces outside this area, it encounters the capacity bottlenecks and has to travel much longer, so there are not many workplaces available at these particular travel times. For this reason, this individual's distribution increases more slowly or even decreases until the first workplace outside the "entrapped area" can be reached. Combinations of configurations like this can lead to plateaus or even local minima in the overall distribution; they reflect typical sizes and distances of isolated (regarding street capacities) regions in the network.

Note that these distributions just reflect the state of the network in the different iterations. Iteration 0 assumes that the network is empty, and thus puts all workplaces within short reach. As a result, iteration 1 is overly congested, and available workplaces are shifted to large times. For all subsequent iterations, traffic gets less and less, which means that the distributions shift back to lower travel times.

Once the working opportunity distribution has been generated based on the last simulation feedback, the travel time acceptance is calculated according to Eq. (4). In Fig. 6 (Middle), the travel time acceptance is shown for different activity iterations.

First, note that in the initial assignment, one clearly sees the effect of the finite problem size: In our empty and finite network, there are simply not enough opportunities for long trips, and the algorithm compensates by putting a really high probability on these few long trips.

After the first workplace re-assignment ("Assignment 1"), the agents prefer much shorter trips. Since, as we know, assignment 1 is an over-reaction to the initial guess, subsequent assignments allow again for somewhat longer trips. Note that the travel time acceptance for workplace assignment number 7 has again the characteristic that long trips are seemingly much preferred – again, the reason for this is that our finite network does not offer enough opportunities for long trips and the algorithm compensates for that. In reality, working opportunities can be found even for much larger travel times under consideration of workplaces in neighbored cities. Two other arguments should be noted in this context, too: (i) People seem to be more indifferent for long travel times than for medium-length travel times [8]. This certainly would not explain why long trip times should be more preferred than medium-duration trip times, but it would make plausible why there could be a smaller slope in the acceptance function for higher travel times. (ii) It is widely believed that the census over-reports travel times; see below. – All these arguments together mean that there is a variety of possible sources of errors for travel times above an hour (3600 min).

At the other end of the travel times, we also obtain lower preferences for very short travel times (below 5 min = 300 sec). This may indicate that people prefer living in a certain minimum distance to their workplace; maybe simply caused by the fact that people would walk really short distances, thus increasing a driving time of for example one minute to a walking time of five minutes. On the other hand, it may also have something to do with the way we smooth our data for short travel times; see Sec. .

Restricting ourselves to the "plausible" range of travel times $400sec < t < 2000sec$, the function $p_{choice}(t)$ can be approximated

$$p_{choice}(t) \sim \exp(\alpha t). \tag{6}$$

Fitting p_{choice} to 6 yields the following values for α:

Re-Assignment	α
1	(-0.0007 ± 0.0003)
5	(-0.0009 ± 0.0003)
7	(-0.0010 ± 0.0003)

Note, and this is really important, that the functional form of the exponential *comes out of the simulations*; it is not invested anywhere in the approach. Thus, what we obtain is another *justification* of ansatz (1), this time not obtained via arguing on the psychological level (as discrete choice theory does), but via making peoples' preferences consistent with their reported travel times in a given transportation network.

Last, we check (Fig. 6 (Bottom)) if the simulated travel time distribution is indeed consistent with the travel time distribution from the census. We see that with higher iteration numbers, the simulated distribution indeed approaches the census distribution, except for too many high travel times. This is somewhat surprising, since in this case one would expect that the acceptance function p_{choice} for "assignment 7" would result in lower weights for long trip times than it actually does. Presumably, the generated travel demands exceed the network capacity in a unrealistic way, and this causes large fluctuations in the travel times of individual travelers [13, 14]. This would mean that always *some* travelers get caught in some heavy traffic, but it is never the same travelers nor the same links, and thus the algorithm in its current form cannot respond to it properly. It is indeed believed that the census overestimates travel times [15]. The explanation for this is that people report the time they *allocate* for the trip, which includes getting ready and walking to the garage, not just the time they are on the road. Also, the simulation model may underestimate network throughput, for example since local streets are missing, and because of the simplifications of the queue model. Resolving this would thus not just mean a precise evaluation (and possible improvement) of our current iteration procedure; it would also involve to calibrate the relation between transportation demand and transportation network throughput. This is beyond the scope of this paper.

Discussion

During the re-assignment iterations we keep household locations fixed and changed working locations. This may be realistic for some people, but for other people it might be proper to argue that instead of looking for a workplace in feasible distance, people move closer to workplaces. In our case, the decision was made by the data that was available: Our demographic information is coupled to household locations, and thus the task is to match people and workplaces given the data, *not* to develop a behaviorally entirely plausible model of what people actually do. Future work will hopefully be able to enhance the behavioral aspect of this work.

Looking at the worker/workplace relations resulting from the stochastic assignment, it is necessary to be aware that different microscopic configurations can lead to the same macroscopic travel time distribution. So far, the assignment procedure is entirely driven by overall workplace availability regardless of detailed information As next step to a more detailed approach it is conceivable to impose additional constraints (e.g. regarding salary) on the assignment procedure. For this, the population as well as the set of available workplaces would be divided into different categories and the stochastic assignment would be applied to each category separately.

Our investigations are confined to home-to-work trips. The underlying dynamics is also applicable for other activity types as far as travel time distributions for those trips are available. Presumably, the acceptance behavior is activity dependent. For example, people accept usually expect longer travel times for their trips to work than for shopping [8]. It would be interesting to see whether the exponential decay in travel time acceptance still holds for other activity types i.e. whether this would only effect the acceptance coefficient α in Eq. 6.

Regarding the iteration sequence, it is possible to combine workplace-assignment and route-planning instead from separating it strictly; i.e. there would be a chance for each individual to change the workplace location in every arbitrary iteration in combination with route-planning. So far, we picked individuals to be re-routed/assigned entirely randomly. It may be possible to speed up the relaxation process by concentrating on "critical" individuals, i.e. individuals that are furthest away from any satisfying choice. For example, one could concentrate on agents where the expected travel time is much different from the travel time experienced in the simulation. In this context, it is also necessary to check in what way the underlying re-routing/assigning algorithm effects the final traffic state after relaxation. Questions related to this are topic of current research.

Summary

We presented an iterative activity assignment approach exemplarily for home-to-work trips. The approach allows to systematically generate macroscopic data – for our investigations this was the census travel time distribution – from a transportation simulation. Instead of making any assumptions about the individuals' travel time acceptance behavior, we extract the acceptance function out of the census travel time distribution and the simulation set-up. For this, we decompose the probability of choosing a workplace in a given distance into two terms: The distribution of workplaces in the network, and the macroscopic travel time acceptance function. It turns out that this acceptance behavior can be well described by an exponentially decaying function, which is consistent with other approaches. The fact that the studies were carried out for the real Portland/Oregon road network shows that iterative activity assignment is feasible for realistically sized systems.

Acknowledgment

We would like to thank R. Frye, R. Jacob, and P. Simon for a lot of useful hints as well as for developing and maintaining the routing tool and the traffic simulator. We are also thankful to C. Barrett and R. Beckman for many fruitful discussions. Finally, we thank P. Wagner, whose visit at LANL gave an essential incentive to focus on this topic.

References

[1] J. Esser and M. Schreckenberg. Microscopic simulation of urban traffic based on cellular-automata. *International Journal of Modern Physics C*, 8(5):1025–1036, 1997.

[2] D. Helbing. *Verkehrsdynamik*. Springer, Heidelberg, Germany, 1997.

[3] B.S. Kerner and P. Konhäuser. Structure and parameters of clusters in traffic flow. *Physical Review E*, 50(1):54, 1994.

[4] R. Barlovic, L. Santen, A. Schadschneider, and M. Schreckenberg. Metastable states in cellular automata. *European Physical Journal B*, 5(3):793–800, 10 1998.

[5] P. Wagner and K. Nagel. Microscopic modeling of travel demand: Approaching the home-to-work problem. Annual Meeting Paper 99 09 19, Transportation Research Board, Washington, D.C., 1998.

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

[7] M. Ben-Akiva and S.R. Lerman. *Discrete choice analysis.* The MIT Press, Cambridge, MA, 1985.

[8] John L. Bowman. *The day activity schedule approach to travel demand analysis.* PhD thesis, Massachusetts Institute of Technology, Boston, MA, 1998.

[9] P.M. Simon and K. Nagel. Simple queueing model applied to the city of Portland. Annual Meeting Paper 99 12 49, Transportation Research Board, Washington, D.C., 1999.

[10] M. Rickert and K. Nagel. Issues of simulation-based route assignment. Los Alamos Unclassified Report (LA-UR) 98-4601, Los Alamos National Laboratory, see `www.santafe.edu/~kai/papers/`, 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] R.J. Beckman, K.A. Baggerly, and M.D. McKay. Creating synthetic base-line populations. *Transportion Research Part A – Policy and Practice*, 30(6):415–429, 1996.

[13] T. Kelly. Driver strategy and traffic system performance. *Physica A*, 235:407, 1997.

[14] B. Raney, J. Esser, M. Hamada, and K. Nagel. In preparation.

[15] K. Lawton. Personal communication.

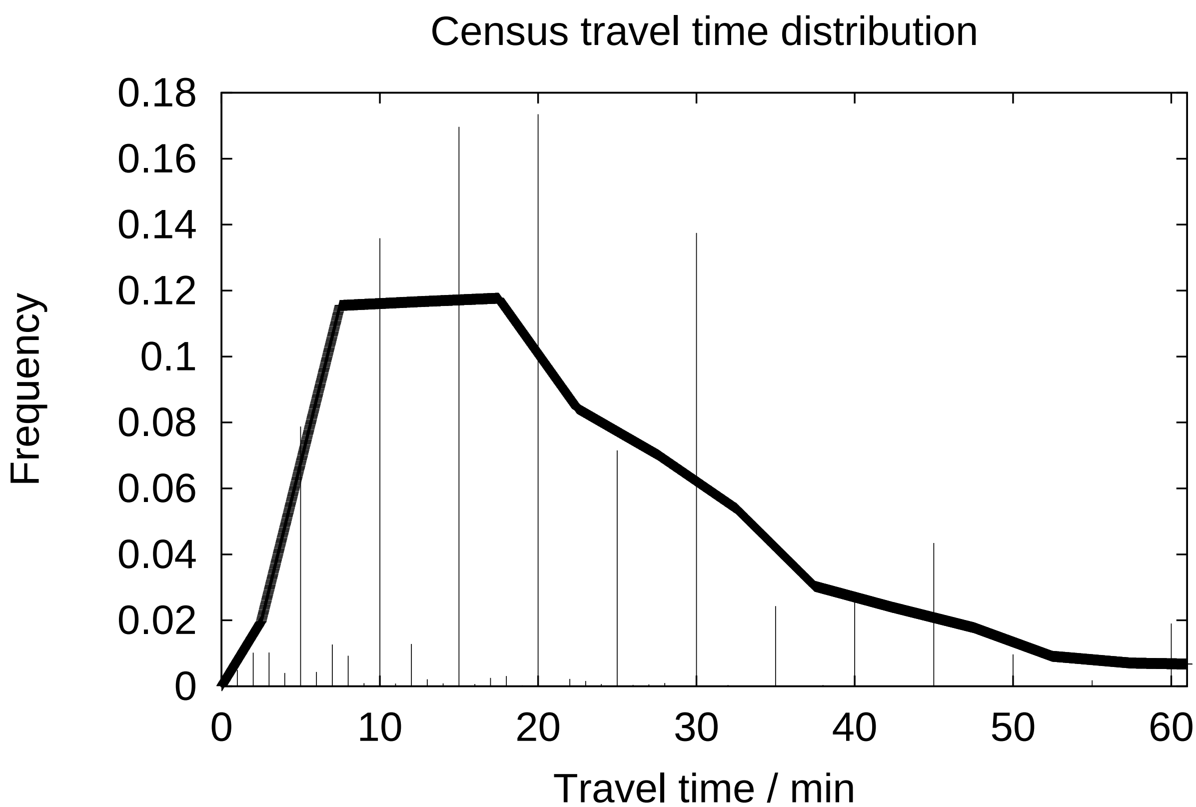


Figure 1: Work trip travel time distribution for home-to-work trips in the Portland study area (Source: Portland Census 1990).

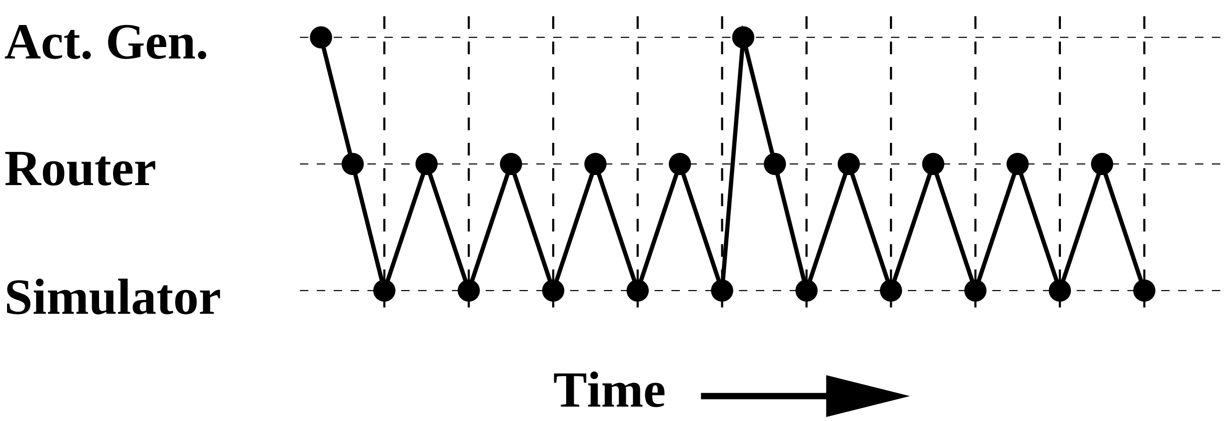


Figure 2: Iterative Activity Re-Assignment: Schematic subsequent application of activity generator, router, and traffic simulator.

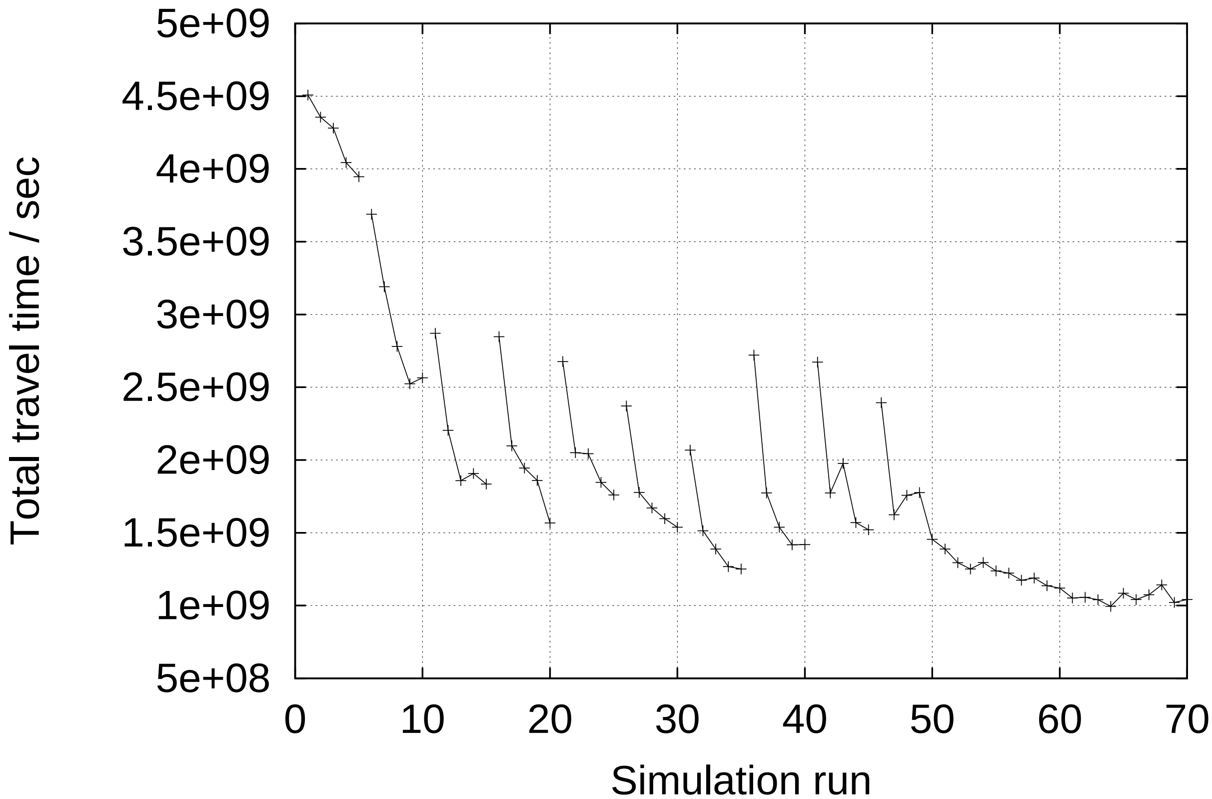


Figure 3: Total travel time in the simulation.

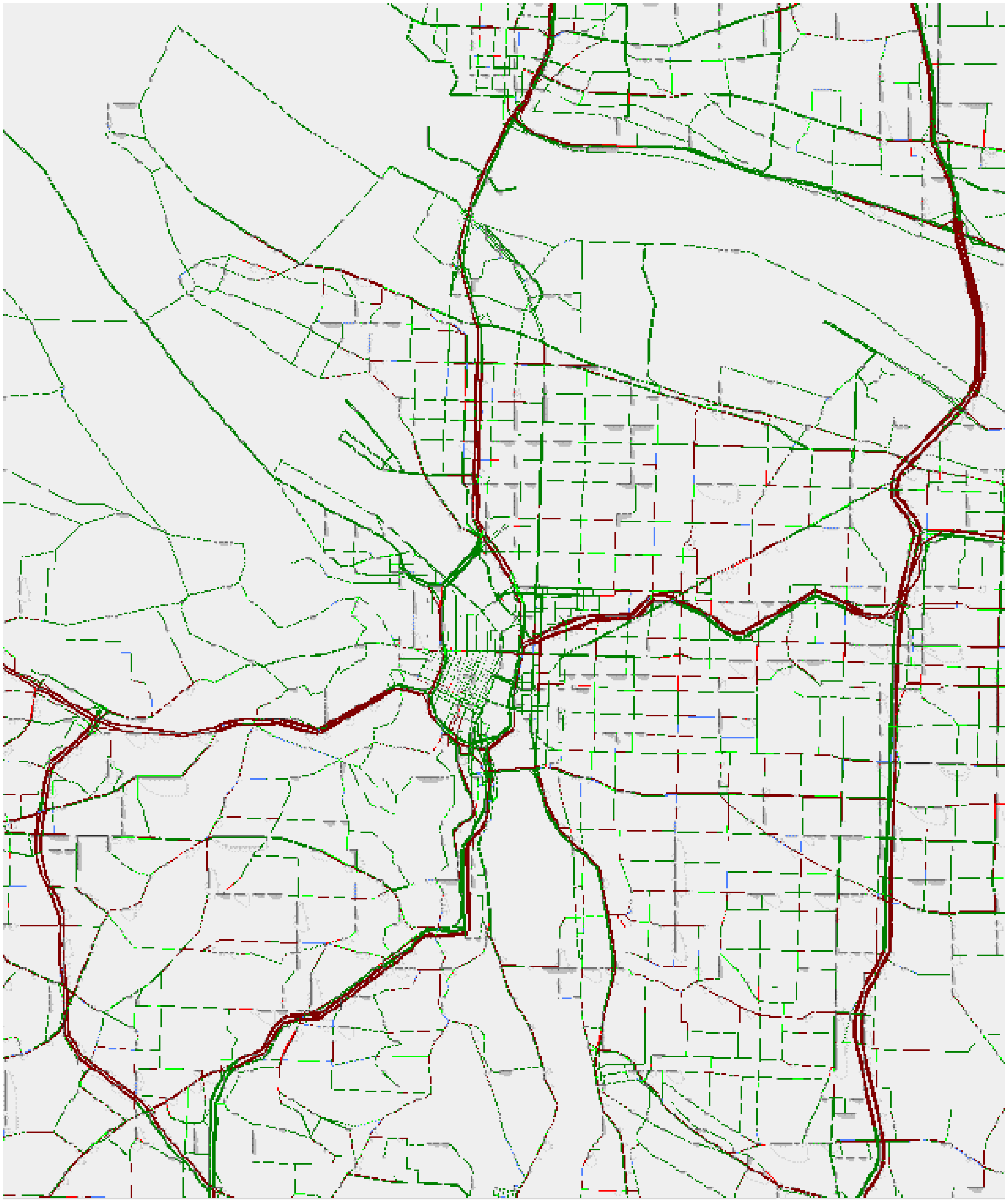


Figure 4: Simulation snapshot at 9am for simulation run 1. Red roads are congested (speed less than 20% of free speed).

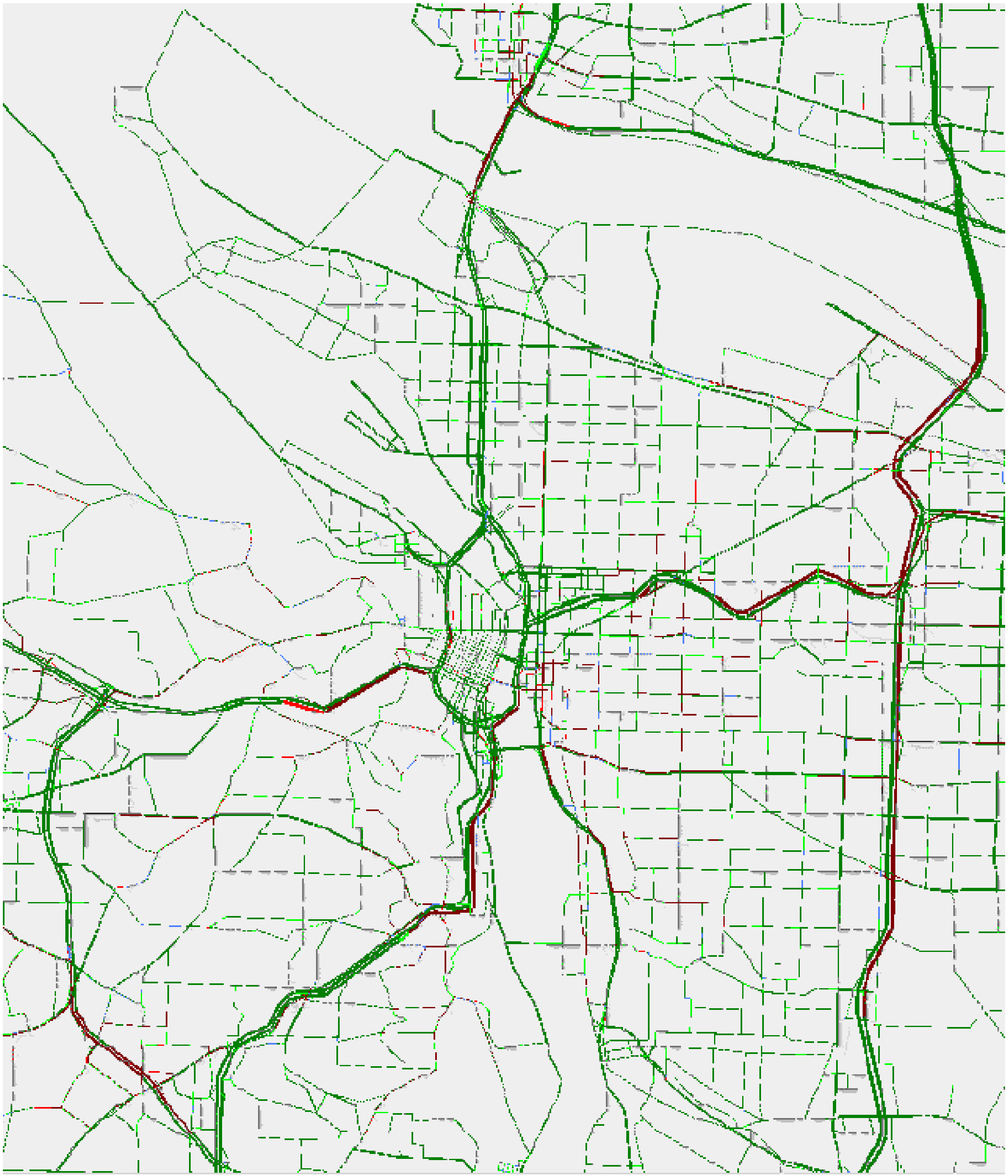


Figure 5: Simulation snapshot at 9am for simulation run 70. Much fewer roads are congested than in Fig. 4.

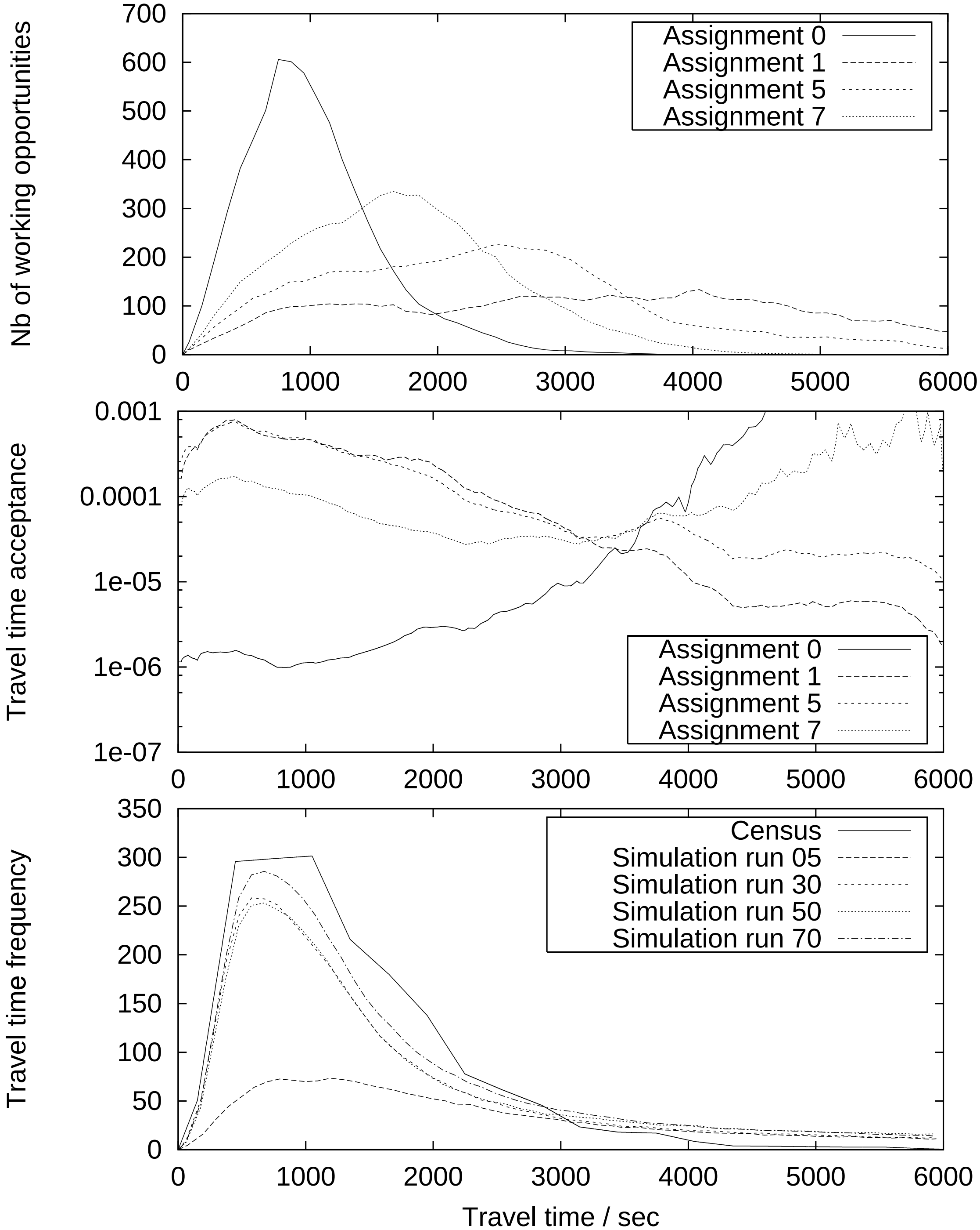


Figure 6: *Top:* Normalized working opportunity distribution. – *Middle:* Travel time acceptance function $p_{choice}(t)$. – *Bottom:* Travel time distribution resulting from the traffic simulator compared to the census distribution.